

قضية مربكة غامضة حان وقت تفكيكها

المعالجة الآلية للغة العربية .. جهود الحاضر وتحديات المستقبل

فى دنيا الحاسبات وتكنولوجيا المعلومات توجد قضايا مربكة وغامضة، لأنها من الناحية العملية والعلمية والمستقبلية تعتبر مهمة وحيوية للغاية وذات تأثيرات عريضة النطاق من النوع الذى يرتبط بمصائر الأمم، لكنها فى الوقت نفسه لا تلقى أى نوع من الجذب الجماهيرى العام لكونها معقدة بدرجة تجعلها أقرب إلى الألغاز من كثرة ما تحمله من تعقيد، ومن ثم فهى تقف وحيدة مع متخصصيها وخبرائها، تعانى قلة الاهتمام والفهم والاستعداد للمساعدة والدعم، وبالنسبة لنا فى مصر والمنطقة العربية تبدو قضية المعالجة الآلية للغة العربية من أبرز هذه القضايا الغامضة المربكة التى لا يلتفت لها الجمهور على الرغم مما تمثله من أهمية قصوى لحاضر ومستقبل لغة الأمة وهويتها ومكانتها فى ظل ثورة المعلومات، إذ لم يعد ممكنا لأى أمة أو ثقافة أن تقف موقف المتفرج على التطور التكنولوجى المتسارع الذى يشهده العالم وإلا حكمت على نفسها إما بالتخلف عن ركب العصر أو فقدت لغتها ومن ثم تراثها، فاللغة العربية - شأن لغات أخرى - تتعرض حاليا لحركة تهميش نشطة بفعل الضغوط الهائلة الناجمة عن طغيان اللغة الإنجليزية على الصعيد السياسى والاقتصادى والتكنولوجى والمعلوماتى، كما تواجه تحديات إضافية نتيجة للحملة الضارية التى تشنها العولمة ضد الإسلام، وبالتالى ضد العربية نظرا لشدة الارتباط بينهما، ولذلك خصصنا ملف هذا الشهر لتفكيك هذه القضية اللغز وشرحها من خلال إلقاء الضوء على جوانبها المختلفة وتبسيطها أمام القارىء، ورصد جهود الباحثين فى مجال المعالجة الآلية للغة العربية المنطوقة والمكتوبة والوقوف على إنجازات معالجة قضايا اللغة العربية وتطويرها وتحديثها حتى تكون أكثر ملاءمة مع تطورات العصر وقادرة على مواجهة تحدياته الكبرى، آملين أن يساعد هذا الجهد فى أن تلقى القضية ما تستحقه من اهتمام ودعم.

ملف إعداد: أبو الحجاج محمد بشير

الخطوة الأولى لفك اللغز

ما معنى المعالجة الآلية للغة ولماذا ظهرت وما تاريخها؟

تفهم الكثرة الغالبة من المستخدمين للحاسبات وتكنولوجيا المعلومات العلاقة بين الحاسب واللغة على أن الحاسب قدم إمكانات كبرى أعانت ـ ولا تزال ـ على دراسة اللغات الإنسانية وتحليلها وتبسيطها وتيسير تعليمها، أما اللغة فكانت ولاتزال هى الوعاء أو الوسيط الذى يتم من خلاله التفاعل بين الإنسان مع الحاسب، بيد أن العلاقة ليست كذلك بالضبط فالخبراء والعلماء والمتخصصون ينظرون إليها نظرة أخرى واضحة الاختلاف، تنطوى على الكثير من التعقيدات والجوانب الغامضة وغير المفهومة بالنسبة للكثيرين، فهم يرون أن القسم الأعظم من هذه العلاقة يقع تحت نطاق المعالجة الآلية للغة، وللوهلة الأولى يبدو المصطلح صعبا وغير مفهوم على الرغم مما يحمله من تحديات جمة ومخاطر عظيمة الشأن خاصة على لغة مثل لغتنا العربية التى تواجه فى هذا السياق تهديدات لا حصر لها، لذلك فإن الخطوة الأولى فى فك اللغز هى أن نعرف ماذا تعنى المعالجة الآلية للغة سواء العربية أو غيرها.

المصطلح والمفهوم

واقعيا تتعدد المصطلحات والتعريفات المعبرة عن هذه قضية المعالجة الآلية للغة، فهناك هندسة اللغة واللغويات الحاسوبية، لكنها تدور جميعا فى دائرة واحدة، وهى تطويع اللغة بكل تعقيداتها وروابطها وشرودها ومجازها لثنائية الصفر والواحد فى برمجيات ونظم الحاسب، ومن ابرز التعريفات السائدة حول معنى المعالجة الآلية للغة التعريف الذى جاء فى موسوعة ويكيبيديا wikipedia.org ومفاده أن المعالجة الآلية للغات الإنسانية Natural Language Processing هى مجال فرعى يتبع الذكاء الاصطناعى واللغويات الحاسوبية، ويعنى بدراسة مشكلات التوليد والفهم الآلى للغات الإنسانية الطبيعية، وتهدف أنظمة توليد اللغات الطبيعية إلى تحويل البيانات والمعلومات المخزنة فى قواعد بيانات الحاسب إلى لغة بشرية تبدو طبيعية، أما أنظمة فهم اللغات الطبيعية فتحويل عينات ونماذج اللغات الإنسانية إلى تمثيل شكلى يسهل على برامج الحاسب تطويعه والتعامل معه.

وقد قدمت جامعة شيفلد البريطانية http://nlp.shef.ac.uk/ شرحا آخر لمعنى المعالجة الآلية للغات الإنسانية قالت فيه أن المعالجة الآلية للغة تعنى استخدام أجهزة الحاسب فى معالجة اللغة المكتوبة والمنطوقة من أجل أغراض عملية مفيدة مثل الترجمة الآلية بين اللغات واستخلاص المعلومات من مواقع الويب وقواعد البيانات وبنوك المعلومات المتصلة بالإنترنت للحصول على إجابات للأسئلة أو من أجل إجراء حوار مع الحاسب أو الآلة للحصول على استشارة أو معلومة ما.

وما سبق مجرد أمثلة فهناك تطبيقات أقل شيوعا ولكنها شيقة ومثيرة للخيال، مثل قدرة الحاسب على أن يحدد إذا كان الخبر المنشور فى صحيفة ما مقتبس من صحيفة أخرى أم لا.

أما اللغويات الحاسوبية Computational Linguistics فيقصد بها الدراسة العلمية للغة من وجهة نظر حاسوبية، حيث يهدف العلماء إلى تقديم نماذج حسابية لأنواع عديدة من الظواهر اللغوية، وهى وفقا لتعريف هانز أوزكوريت، أستاذ اللغويات الحاسوبية فى جامعة سارلاند الألمانية، علم يقع فى مرتبة وسيطة بين اللغويات وعلوم الحاسب، التى تهتم بالجوانب الحاسوبية لملكة اللغة البشرية، وينتمى هذا العلم إلى فئة العلوم الإدراكية ويتداخل مع الذكاء الاصطناعي، وهى فرع من علوم الحاسب التى تهدف إلى تقدم نماذج حوسبية للإدراك البشري.

يصف **الدكتور نبيل على** ـ خبير المعلوماتية والمفكر العربى الكبير ـ فى كتاب (الثقافة العربية وعصر المعلومات، عالم المعرفة، ٢٠٠١) علاقة اللغة بالحاسب بأنها علاقة منفعة متبادلة، فعلى جبهة اللغة يستخدم الحاسب حاليا لإقامة النماذج اللغوية وتحليل الفروع اللغوية المختلفة، ومن أمثلة تطبيقات الحاسب فى مجال اللغويات الصرف الحاسوبى والنحو الحاسوبى والدلالة الحاسوبية والمعجمية الحاسوبية وعلم النفس اللغوى الحاسوبي. وفى المقابل اقتبس علماء الحاسب فى تطويرهم للغات البرمجة الكثير من أسس اللغات الطبيعية ويسعون بخطى حثيثة إلى التقريب بين هذه اللغات الاصطناعية واللغات الطبيعية بهدف تسهيل التعامل مع الحاسب دون وسيط برمجي، فالهدف الأسمى لبرمجة الحاسب هو أن يتعامل الفرد معه مباشرة بلغته الطبيعية بدلا من اللغات الاصطناعية.

وحتى لا نستفيض فى تعريف اللغز سنكتفى بهذه المصطلحات والتعريفات القليلة، ونخلص منها إلى أن المعالجة الآلية للغة هى جهود تحاول إزالة الحواجز ما بين اللغة التى يستخدمها الإنسان العادى فى ظروفه الطبيعية، والحاسب كآلة ذات قدرات عالية فى فهرسة وتخزين ومعالجة واستدعاء البيانات والمعلومات، بما يجعل الإنسان قادر على استثمار أقصى طاقات وإمكانات الحاسب بسهولة ويسر وعبر لغة التعامل الطبيعية، وبما يجعل الحاسب قادر على أن يفهم لغة الإنسان العادية الطبيعية على مستوى الكلمة والجملة والمعنى وينفذ ما يريده الإنسان عبر هذا الفهم.



الأهداف والأسباب

ولكن لماذا المعالجة الآلية للغات؟ يمكن القول أن المعالجة الآلية للغة ظهرت فى الأساس من أجل تطوير وإنشاء وصياغة تطبيقات الترجمة الآلية سواء كانت مكتوبة أو منطوقة، والترجمة الآلية هى الوسيلة الوحيدة التى تستطيع بها الأمم مواكبة الانفجار المعرفى والمعلوماتى فى عصر الإنترنت الساحق وفضاءاتها الإلكترونية الواسعة. حيث لا يستطيع أى مجتمع أو أمة مهما كانت أن تعيش بمعزل عن هذه الثورة المعرفية أمام اكتساح اللغة الإنجليزية لكل المراجع المعرفية والمعلوماتية فى فضاء الإنترنت وانتشارها بشكل يهدد بانسحاق كل الأمم واللغات التى لا تستخدم تكنولوجيا المعلومات والحاسب فى اللحاق بعصر المعلومات.

ومن أحد أهداف وتطبيقات المعالجة الآلية للغة مساعدة المستخدم والإدارى والعالم والمسئول الحكومى فى عصر فيض وطغيان المعلومات على الوصول للمعلومات التى يريدها بسهولة وسرعة من خلال التلخيص الآلى والتحليل الآلى للنصوص بالإضافة إلى البحث الذكى عن المعلومات فى شبكة الإنترنت فى أجيالها الجديدة التى تعرف باسم الويب الدلالية.

ويؤدى التطور فى تطبيقات المعالجة الآلية للغة إلى تسهيل حصول المرء على المعلومة فى أى مكان وأى وقت، ومن أمثلة السيناريوهات المستقبلية لتوضيح هذا الأمر تخيل نفسك أنك تقود سيارتك فى رحلة لقضاء عطلة الصيف فى الساحل الشمالي، وكنت تريد متابعة أسهم شركتك فى البورصة، فماذا ستفعل؟ ستخاطب حاسب السيارة قائلا: أريد تقريرا تفصيليا عن حالة أسهم شركة كذا فى البورصة خلال اليومين الماضيين؟ ماذا الذى سيحدث؟ سيقوم حاسب السيارة بتحويل عبارتك الصوتية إلى أوامر بحث نصية ويرسلها إلى شبكة الإنترنت عبر التقنيات اللاسلكية

المتقدمة، وستقوم شبكة الإنترنت بفحص أسهم شركتك خلال اليومين السابقين وترسلها إلى حاسب السيارة مرة أخرى الذى يحول المعلومات النصية إلى كلام منطوق تسمعه بأذنيك لأنك بطبيعة الحال تركز فى القيادة.

ويمكن اختصار أهداف المعالجة الآلية للغات الإنسانية بصفة عامة فى ثلاثة أهداف هى:

١- تواصل أفضل مع الحاسب

تمكن وسائل البحث باللغة العادية الإنسان من التواصل مع الحاسب بالفرنسية أو الألمانية أو العربية أو أى لغة أخرى. فالتواصل مع الحاسب باللغة المنطوقة سيكون له تأثير كبير على بيئة العمل، وستنبثق مجالات جديدة واسعة أمام تكنولوجيا المعلومات.

٢- تواصل أفضل بين البشر

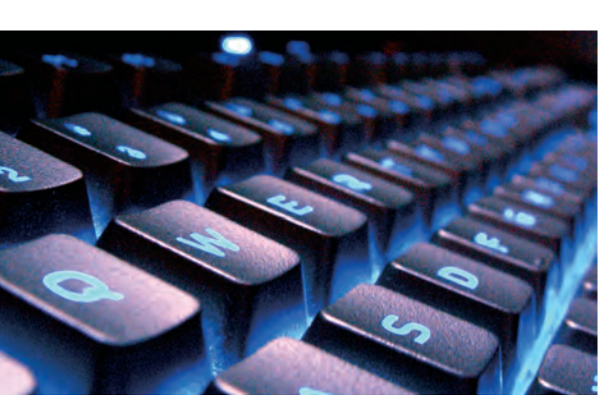
من الأهداف الأولى التى ظهرت من أجلها علوم اللغويات الحاسوبية والمعالجة الآلية الترجمة الآلية بين اللغات الحية من أجل مزيد من التواصل بين البشر، ورغم أن تجربة الفشل المريرة قد جعلت العلماء يدركون أنهم بعيدون جدا عن تحقيق هذا الهدف الطموح فى ترجمة النصوص غير المحدودة، تمكن علماء اللغويات الحاسوبية من إنشاء برمجيات تبسط عمل المترجم البشرى وتحسن من إنتاجيته إلى حد كبير، وتقدم الترجمة الآلية الركيكة أو الحرفية مساعدة كبيرة لباحثى المعلومات الذين يبحثون عن المدلول أو المعنى فى كميات كبيرة من النصوص باللغات الأجنبية.

٣- الوصول الفعال للمعلومات

تصفح المعلومات على الويب والتنقل بينها وفلترتها ومعالجتها يتطلب تطوير برمجيات يمكنها الوصول إلى المعلومات فى المستندات والوثائق وصفحات الويب، وهكذا فإن تكنولوجيا اللغات الإنسانية لإدارة المحتوى شرط ضرورى لتحويل ثروة المعلومات الرقمية إلى معرفة جماعية، وتعدد لغات المحتوى على الويب يمثل تحدياً إضافياً لتكنولوجيا اللغة، لأنه لا يمكن السيطرة على الويب العالمية إلا بمساعدة الأدوات متعددة اللغات لفهرسة وتصفح الويب، وستذلل أنظمة إدارة المعرفة والمعلومات متعددة اللغات عقبات اللغة أمام التجارة الإلكترونية والتعليم عن بعد.

البداية والتاريخ

لم تظهر المعالجة الآلية للغة فجأة بل كانت جهودا متواصلة متراكمة حققت تقدما لا بأس به حتى الآن، وإذا ما استعرضنا تاريخ هذه القضية المعقدة سنجده يعود إلى أربعينيات القرن الماضى مع أول ظهور للحاسبات، فساعتها ساد الشعور بالتفاؤل فى الأوساط السياسية والعسكرية تجاه إمكانية استخدام قدراته فى التحليل اللغوى والترجمة الآلية، ففى منتصف الخمسينيات اتجهت الجهود العسكرية فى الولايات المتحدة إلى استخدام الحاسب فى الترجمة الآلية للنصوص من اللغات الأجنبية وخاصة الدوريات العلمية الروسية إلى الإنجليزية. ولأن الحاسبات قد أثبتت قدرتها الفذة على القيام بالعمليات الحسابية بسرعة أكبر ودقة أعلى من الإنسان، كان هناك اعتقاد شائع بأنه لن يمضى وقت طويل قبل أن تتمكن من معالجة اللغة بنفس كفاءة العقل البشري، وأن الأمر سهل للغاية من خلال إنشاء قواميس آلية ثنائية اللغة، وسرعان ما باءت المحاولات الأولى بفشل ذريع، لأن الحاسب أو أنظمة تشغيله أو لغات برمجته وقتها لم تكن قد بلغت من النضج والتطور ما يؤهلها لدراسة وتحليل تعقيد اللغة وغموضها ومجازها، لأن ذلك لا يتطلب فهم معانى الألفاظ فحسب، بل يتطلب كذلك معرفة عميقة بالعالم الخارجي مع القدرة على استغلال هذه المعرفة بشكل ذكي، واعترف العلماء بصعوبة إخضاع اللغات للمعالجة الآلية قبل أن تتوفر القاعدة المعرفية الكافية، وهكذا ولدت علوم ودراسات اللغويات الحاسوبية ومعالجة اللغات بواسطة الحاسب.



وبعد أن تطورت النظريات الإحصائية والرياضية لدراسة اللغة، كان من المنطقى بل والحتمي أن تأتى المرحلة التالية وهى التقاء الحاسب باللغة، لأن اللغة تجسد النشاط الذهنى للإنسان بكل تجلياته، ولأن هدف الحاسب هو محاكاة القدرات الذهنية للإنسان، وقد تدرج هذا الالتقاء بين الحاسب واللغة حتى وصل إلى درجة عالية من التفاعل والامتزاج لعدة أسباب ذكرها الدكتور نبيل على فى كتابه (العرب وعصر المعلومات، عالم المعرفة، ١٩٩٤) على النحو التالي:

● التطور الهائل فى علوم اللغويات وخضوعها للمعالجة الرياضية والمنطقية.

● التقدم العلمى فى تقنيات الحاسب ومكوناته ولغات برمجته وأساليب الذكاء الصناعى.

● الاستعانة باللغات البشرية فى تصميم لغات برمجة راقية تتسم بالقوة والمرونة.

● يتطلب الانفجار المعرفى استحداث وسائل آلية لتنظيم هذا الكم المتزايد من المعلومات وتحسين كفاءة تخزينها واسترجاعها وتوظيفها.

● انتشار الحاسبات الشخصية والمنزلية بشكل يحتم ضرورة التعامل معها بلغة طبيعية .

● بفضل الحاسبات السوبر أمكن تطوير نظم لمعالجة اللغة آليا فى حدود الجدوى الاقتصادية والفنية لهذه النظم.

● ظهور النظم الآلية الخبيرة التى تستطيع تشخيص الأمراض وتقديم الاستشارات الفنية والقانونية والنظم الآلية للتعليم الذاتي، التى تتطلب قدرة على الحوار مع المستخدم البشرى بلغة سهلة مثل لغته الطبيعية.

●انتشار الحاسب كوسيلة للتعليم والتعلم، وخاصة تعليم وتعلم اللغات.

تحديات تواجه المعالجة الآلية للغة

فهم الحاسب للغات الطبيعية من المشكلات الكبرى التى تواجه نظم المعالجة الآلية للغات الطبيعية، لأن ذلك يتطلب معرفة عميقة بالعالم الخارجي مع القدرة على استغلال هذه المعرفة بشكل ذكي، وفيما يلى نماذج للمشكلات التى تعترض المعالجة الآلية للغات بواسطة الحاسب، ومن ضمنها بالطبع اللغة العربية:

١-تقطيع الكلام والأصوات والوحدات المعجمية

فى أغلب اللغات المنطوقة، تعد الكلمات توليفة من الأصوات المتتالية الممتزجة مع بعضها البعض، ففى الكلام العادى الطبيعي، نادرًا ما توجد أى وقفات بين الكلمات المتتالية ويتأثر الحرف بالحروف المجاورة من عدة نواحي، حيث تمتزج الأصوات ببعضها بشكل ناعم وسلس أو تنفصل أو حتى تكاد تختفي. لذلك يمثل تقطيع الجملة إلى كلمات منفصلة وتحويل الكلمات إلى حروف متفرقة بل وتقطيع الحروف داخل الكلمة مهمة شديدة الصعوبة فى تكنولوجيا التعرف على الكلام .

وعلاوة على ما سبق، من الممكن أن يختلف معنى أى جملة حسب طريقة تقسيمها إلى كلمات، والتقطيع المعجمى السليم يعتمد على السياق والدلالة، وتتداخل هذه المشكلة إلى حد ما مع مشكلة تقطيع النص فى بعض اللغات التى تكتب دون فواصل بين الكلمات مثل الصينية واليابانية واللغات التى تكتب بحروف متصلة مثل اللغة العربية والفارسية والأردية.

٢- فك الغموض أو اللبس

فى كل اللغات توجد بعض الكلمات التى تتعدد معانيها ويتحدد معناها حسب موقعها فى الجملة أو السياق، وتعد مشكلة الغموض أو اللبس فى المعنى وتعرف باسم اللبس الدلالى من أكبر التحديات التى تواجه معالجة اللغة العربية آليا وتطبيقاتها وخاصة الترجمة الآلية، بجانب مشكلة الضمائر التى قد تعود على عاقل أو غير عاقل بعكس الإنجليزية التى تفرق بين العاقل وغير العاقل فى الضمائر سواء كانت ضمائر الفاعل أم المفعول أم الملكية. وهناك غموض آخر بسبب بناء أو تركيب الجملة يسمى اللبس النحوي، حيث تعنى الجملة أكثر من معنى أو يمكن تفسيرها بأكثر من طريقة مثل (شاعر النيل العظيم) حيث يمكن أن تعود صفة العظيم على النيل أو الشاعر.

٣- العبارات الطلبية

بعض الجمل لا تعنى ما تقوله حرفيا، ويتحدد معناها على الموقف الاجتماعي، كأن تقول لشخص على سبيل المثال: هل من الممكن أن تعطينى الملح بجوارك؟ فهذا ليس سؤالا يحتاج إلى إجابة، ولكنه طلب، وكل أشباه هذه العبارات التى تعنى غير مدلولها الحرفى تمثل مشكلات وتحديات شديدة الصعوبة أمام عقل الحاسب الذى يصعب أن يفهم المواقف الاجتماعية أو يفسرها أو يتصرف وفقا لها.

تشمل الترجمة والتوليد والتلخيص الآلى للغة وفهم الصوت

١٠ تطبيقات حالية ومتوقعة للمعالجة الآلية للغة

عبر رحلتها التاريخية وما أنجزته خلالها من نجاحات وتطورات متنوعة استطاعت المعالجة الآلية أن تحفر لنفسها مجالات للتطبيق العملى فى مسارات متنوعة، بعضها بدأنا نلمس آثاره عمليا مثل أدوات الترجمة الآلية المتوفرة حاليا مجانا على محركات البحث الشهيرة مثل جوجل، وبعضها الآخر لايزال فى البداية ولم يحقق الانتشار والشعبية المطلوبة، وفيما يلى نتناول بعض تطبيقات المعالجة الآلية للغة والتى تشمل الترجمة الآلية والتلخيص الآلى والتوليد الآلى للغة واستخلاص المعلومات واسترجاع المعلومات والإجابة على الأسئلة والتنقيب فى النصوص وتحويل النص إلى كلام منطوق وفهم الصوت.

١- الترجمة الآلية

وتعنى استخدام برمجيات الحاسب فى ترجمة النصوص أو الكلام من لغة إنسانية لأخرى. وفى مستواها الأساسى، تعمل برامج الترجمة الآلية على استبدال الكلمات باللغة المترجم منها بالكلمات المقابلة لها فى اللغة المترجم إليها. من الممكن استخدام تقنيات المكانز فى إجراء عمليات ترجمة أكثر تعقيدا حيث تساهم المكانز والذخائر اللغوية فى التعامل مع الفروق فى البنية اللغوية والتعرف على العبارات وترجمة المصطلحات بالإضافة إلى عزل الحالات الشاذة.

تتيح برمجيات الترجمة الآلية الحالية تخصيص الترجمة حسب المجال أو المهنة، حيث يتم تحسين الترجمة النهائية من خلال حصر نطاق الاستبدالات المسموح بها، وهذا الأسلوب فعال للغاية خاصة فى المجالات التى تستخدم فيها اللغة الرسمية أو الاصطلاحية، وبصفة عامة تتيح الترجمة الآلية نتائج أفضل فى النصوص الحكومية والقانونية التى تعتمد على قوالب من الجمل والعبارات بعكس النصوص العامة والمحادثات، حيث ما زالت نظم الترجمة الآلية فى حاجة إلى مزيد من التطوير للوصول إلى جودة معقولة.

ومما لا شك فيه أن الترجمة الآلية بصفة عامة وصلت إلى مستوى متقدم جدا، ونجحت فى مساعدة المترجم البشرى على تحسين عمله بل وتفوقت عليه فى أحيان قليلة، غير أنها فى مجملها لم تستطع أن تتغلب على المترجم البشرى فى هذا المضمار، خاصة فى ترجمة الحوارات والمحادثات ويزيد الأمر صعوبة لو كانت هذه الحوارات باللغة العامية أو اللغة غير الرسمية.

وتستخدم الترجمة الآلية طريقة تعتمد على القواعد اللغوية التى تعنى ترجمة الكلمات بطريقة لغوية، حيث يتم استبدال الكلمات المناسبة فى اللغة الهدف بالكلمات المقابلة لها فى اللغة المصدر، وهناك بعض الآراء التى ترى أنه لن يكتب للترجمة الآلية تحقيق النجاح ما لم تحل مشكلة فهم اللغات الطبيعية أولا.

وتوجد عدة طرق للترجمة الآلية هي: الترجمة المعتمدة على القواعد أو المعتمدة على الإحصاء. تقوم لوغاريتمات الترجمة المعتمدة على القواعد بإعراب النص وإنشاء تمثيل وسيط رمزي، يتم منه توليد النص باللغة الهدف، ووفقا لطبيعة التمثيل الرمزي، يوصف منهج الترجمة الآلية بأنه معتمد على التمثيل المحايد أو التحويل. وتتطلب هذه المناهج معاجم شاملة بمعلومات دلالية وتركيبية وصرفية ومجموعة كبيرة من القواعد.

وتحاول الترجمة الآلية الإحصائية توليد الترجمة باستخدام الطرق الإحصائية المعتمدة على ذخيرة لغوية ثنائية اللغة، فإذا توافرت هذه المكانز، يمكن تحقيق جودة ممتازة فى الترجمة الآلية لأى نصوص مشابهة، واول برنامج للترجمة الإحصائية هو CANDIDE من آى بى إم، وقد استخدمت شركة جوجل SYSTRAN لعدة سنوات ثم انتقلت إلى طريقة الترجمة الإحصائية فى أكتوبر ٢٠٠٧، وقامت جوجل مؤخرا بإضافة ٢٠٠ مليار كلمة من مواد الأمم المتحدة لتدريب أنظمة الترجمة الآلية، حيث تحسنت جودة الترجمة كثيرا.



٢- التلخيص الآلى

تطبيق من تطبيقات المعالجة الآلية للغة يقوم إنشاء نص مختصر من ملف أو مستند بواسطة برنامج حاسب آلي، على أن يحتوى النص المختصر على أهم الأفكار فى النص الأصلي، وتأتى أهمية التلخيص الآلى فى ضوء إغراق المعلومات وزيادتها عن قدرة المرء على الملاحقة والمتابعة.

ينبغى على البرمجيات التى طورت لتقدم خلاصات متماسكة أن تأخذ فى الاعتبار عدة متغيرات مثل الطول وأسلوب الكتابة والبناء من أجل إنشاء ملخص مفيد.

ويمكن التمييز بين نوعين من برامج التلخيص الآلي: برامج الاستخلاص والتجريد. تعتمد برمجيات الاستخلاص على نسخ المعلومات التى تعد مهمة إلى الملخص (مثل الجمل الأساسية والفقرات المهمة) أما التجريد أو التركيز فيتطلب إعادة الصياغة، وبصفة عامة، يعد التجريد وإعادة الصياغة أقوى تأثيرا ويركز المعلومات بصورة أكبر من الاستخراج، ولكن البرامج التى تقوم بذلك صعبة للغاية فى البرمجة والتطوير لأنها تتطلب تكنولوجيا توليد اللغة الطبيعية، التى ما زالت حتى الآن مجالا متناميا.

٣- التوليد الآلى للغة

يقصد بالتوليد الآلى اللغة إنشاء نص بلغة طبيعية من نظام تمثيل آلى مثل قاعدة معرفة أو استمارة منطقية، وبعض الناس يعتبر التوليد الآلى للغة

كمقابل لفهم اللغة الطبيعية، وفى نظم إنشاء التوليد الآلى للغة، يحتاج النظام إلى اتخاذ قرارات بشأن كيفية صياغة أحد المفاهيم. وأنجح التطبيقات للتوليد الآلى للغة أنظمة تحويل البيانات إلى نصوص التى تقوم بإعداد خلاصات نصية للبيانات الرقمية وغير اللغوية حيث تمزج بين تحليل البيانات والتوليد الآلى للغة مثل النشرات الآلية لأحوال المناخ والاقتصاد والبورصة.

٤- استخلاص المعلومات

يقصد باستخلاص المعلومات فى معالجة اللغات الطبيعية استرجاع المعلومات، فالهدف هو الاستخلاص الآلى للمعلومات المنظمة المصنفة والمعرفة من حيث السياق والمعنى من مستندات إلكترونية غير منظمة. ومن الأهداف الواسعة لاستخلاص المعلومات استنتاج الاستدلالات من المحتوى المنطقى للبيانات المدخلة، ويأتى هذا الهدف فى ظل نمو المعلومات فى الأشكال غير المنظمة (أى دون بيانات وصف) على الإنترنت، والتى يمكن تسهيل الوصول إليها من خلال ترميزها بأكواد XML. ومن التطبيقات المثالية لاستخلاص المعلومات مسح مجموعة من المستندات المكتوبة بلغة طبيعية وحشو قاعدة البيانات بالمعلومات التى تم التوصل إليها. ومن المهام الفرعية لاستخلاص المعلومات: تمييز أسماء الأعلام تلقائيا وتمييز العبارات الاسمية التى تشير إلى نفس الشيء واستخلاص المصطلحات أى العثور على المصطلحات لمكنز لغوى معين واستخلاص العلاقات بين الأشياء أو الهويات.

٥- استرجاع المعلومات

خرجت تطبيقات استرجاع المعلومات من عباءة علم البحث عن الوثائق والمستندات والمعلومات داخل الحاسبات وقواعد ومستودعات البيانات وشبكة الويب العالمية وغيرها، وهناك تداخل فى استخدام هذا المصطلح لاسترجاع البيانات والمستندات والمعلومات والنصوص، واسترجاع المعلومات كعلم متعدد الفروع وتتضافر فيه علوم الحاسب والرياضيات وعلوم المكتبة وعلوم المعلومات وبنية المعلومات وعلم النفس الإدراكى واللغويات والإحصاء والفيزياء. تستخدم أنظمة استرجاع المعلومات آليا فى تقليل فيض المعرفة، فالعديد من الجامعات والمكتبات العامة تستخدم أنظمة استرجاع المعلومات لتسهيل العثور على الكتب والدوريات وغيرها من الوثائق ومحركات البحث هى التجسيد الأسمى لتطبيقات استرجاع المعلومات. تتمثل علاقة استرجاع المعلومات بمعالجة اللغة آليا فى تطوير قدرات محرك البحث أو قاعدة البيانات على فهم العبارات والأسئلة باللغة العادية مثل (أريد كتابا يتحدث عن الفيزياء الفلكية) وفى ربط أنظمة استرجاع المعلومات بأنظمة تحويل الكلام إلى نص وبتحويل النص إلى كلام منطوق.

٦- الإجابة على الأسئلة

يعتبر هذا التطبيق من التطبيقات المنتمية استرجاع المعلومات، حيث يفترض بالحاسب أن يستطيع الإجابة على أى أسئلة باللغة العادية من خلال البحث فى مجموعة ضخمة من المستندات والوثائق مثل شبكة الويب العالمية، وتتطلب الإجابة على الأسئلة وسائل معالجة آلية متقدمة للغات مثل استرجاع المستندات، ويعتبرها الكثيرون المرحلة التالية بعد محركات البحث، حيث تقوم بطرح سؤال عادى عن شيء معين، وينبغى على الحاسب أو محرك البحث أن يعرض لك الإجابة فقط أو المستندات التى تحوى الإجابة فقط، وبعض الخبراء يقولون إن الإجابة على الأسئلة لن تكون ممكنة إلا بعد تطوير الويب الدلالية خلال السنوات الخمس أو العشر المقبلة، وحتى الآن ما زالت الإجابة على الأسئلة المنطوقة شكل من أشكال الخيال العلمى حيث يسأل الكابتن السفينة الفضائية عن أى شيء وتجيبه بصوت أنثوى رصين.

٧- التنقيب فى النصوص

يقصد به عملية استخلاص معلومات عالية الجودة من النصوص، وتستمد المعلومات عالية الجودة من تقسيم الأنماط والاتجاهات من خلال وسائل مثل التعلم الإحصائى للأنماط. وتتضمن عملية التنقيب فى النصوص هيكلة النصوص المدخلة من خلال الإعراب (الفك إلى الوحدات اللغوية) مع إضافة مزايا لغوية مشتقة وإزالة مزايا أخرى والإدخال التالى فى قاعدة البيانات واشتقاق الأنماط داخل البيانات المهيكلة وفى النهاية تقييم وتفسير المخرجات. تتضمن مهام التنقيب فى النصوص تصنيف النصوص وعنقدتها واستخراج المفاهيم والهويات وإنتاج التصنيفات المتدرجة وغيرها. وتستخدم تقنيات التنقيب فى النصوص فى تطبيقات الحماية والرعاية الطبية والبرمجيات والتطبيقات وتحسين نتائج البحث وأغراض التسويق والتطبيقات الأكاديمية.

٨- تحويل النص إلى كلام منطوق

من التطبيقات المهمة للمعالجة الآلية للغة، فهو يقوم بقراءة النصوص أو تحويل الكلام المكتوب إلى صوت مسموع وكلام منطوق مفهوم، ويسمى نظام الحاسب (برمجيات أو أجهزة) المستخدم لهذا الغرض بمولف الكلام، ويقوم نظام تحويل النص إلى كلام بتحويل نص اللغة العادية إلى كلام، أما الأنظمة الأخرى فتعمل على تحويل الرموز اللغوية الصوتية إلى كلام. يمكن إنشاء الكلام المولف من خلال ضم أجزاء متسلسلة من الحديث المسجل المخزن فى قاعدة بيانات هى المكانز المنطوقة، وتختلف الأنظة فى حجم وحدات الحديث المخزنة، وفى مجالات استخدام معينة، يتيح تخزين كلمات كاملة أو جمل كاملة إنتاج كلام عالى الجودة. وهناك طريقة أخرى هى تضمين نموذج من جهاز النطق وغيرها من خصائص الصوت الإنسانى لإنتاج صوت مولف بالكامل. ويتم الحكم على جودة مولف الكلام بدرجة تماثله مع الصوت البشرى أو بمدى فهمه، ويتيح برنامج تحويل النصوص إلى كلام مفهوم للمكفوفين والمعاقين بصريا الاستماع إلى الأعمال المكتوبة من خلال الحاسب المنزلي، وقد تضمنت العديد من أنظمة تشغيل الحاسب مولفات كلام منذ بداية الثمانينات. المشكلة التى تواجه إنتاج مولفات كلام باللغة العربية هى غياب علامات التشكيل، ولهذا لابد من إنتاج المشكل الآلى حتى يمكن تحويل النصوص العربية إلى كلام منطوق.

٩- فهم الصوت

يختلف هذا التطبيق speech recognition عن السابق فى أنه لا يتعرف على نص مكتوب ويحوله إلى كلام منطوق، بل يستمع إلى صوت مسموع ويقوم بالتعرف عليه وتحديد هوية صاحبه، من خلال تحويل الصوت إلى رموز تفهمها الآلة وتتعرف عليه، وهو يختلف كذلك عن مصطلح أما مصطلح التعرف على الصوت voice recognition فيقصد به التعرف على المتحدث نفسه وليس الكلام الذى يقوله. ومن تطبيقات فهم الكلام الاتصالات الصوتية وتوجيه المكالمات والتحكم فى الأجهزة المنزلية والبحث فى المحتوى بالصوت وإدخال البيانات البسيطة وإعداد المستندات المنظمة وتحويل الكلام إلى نص مكتوب وفى كابينات القيادة بالطائرات.

ومن تطبيقات فهم الكلام المنطوق استخدامها فى القيادة الآلية للطائرات العسكرية والتحكم فى أجهزتها وخاصة فى بريطانيا وفرنسا والولايات المتحدة وفى طائرات الهليكوبتر حيث مشكلة الضوضاء الخلفية بسبب صوت المروحة والهواء وفى إدارة المعارك حيث تتطلب مراكز القيادة الوصول السريع لقواعد بيانات المعلومات المتغيرة بسرعة كما تستخدم فى تدريب مراقبى حركة المرور الجوي، بجانب استخدامها فى مجال الاتصالات التليفونية وألعاب الحاسب والمحاكاة، ولم يتم تثبيت هذه التكنولوجيا فى الأجهزة المحمولة لأنها تتطلب قوة معالجة هائلة، وهذه التكنولوجيا مفيدة للغاية لمن لا يستطيعون تحريك أيديهم، مما يتطلب وسيلة بديلة لإدخال المعلومات فى الحاسب والتحكم فى وظائفه.

١٠-التعرف الضوئى على الحروف

يقصد بالتعرف الضوئى على الحروف OCR التحويل الميكانيكى أو الإلكترونى لصور الكتابة اليدوية أو الكتابة بالآلة الكاتبة أو النص المطبوع، والتى يتم عادة التقاطها بالماسحة الضوئية إلى نص قابل للتحرير والقراءة فى الحاسب. وقد بلغت تقنية التعرف الضوئى على الحروف شأوا كبيرا فى اللغات اللاتينية بل لم تعد مشكلة على الإطلاق، وفى اللغة العربية توجد تطبيقات متطورة للتعرف الضوئى على الحروف العربية المطبوعة من صخر غير أنها باهظة التكلفة وقليلة الانتشار. وما زال التعرف الضوئى على الكتابة باليد أو الكتابة المنحنية المتصلة مجالا للبحث النشط سواء فى اللغات اللاتينية أو اللغة العربية.



تعمل كبيت للذاكرة ونبع متدفق للمعرفة

المكانز الإلكترونية القلب النابض للمعالجة الآلية للغات

لم يكن ممكنا للمعالجة الآلية للغات أن تتقدم وتفتح لنفسها مجالات التطبيق المتنوعة السابقة ثم تستشرف المستقبل بدون أن يكون لها قلب نابض يستوعب مليارات الكلمات والأصوات، وتكون لديه القدرة على الوصول إلى أى منها وفهم وتحليل ما بينها من علاقات، ليصبح مع الوقت بيتا للذاكرة ونبع متدفق للمعرفة، وعمليا تجسد هذا القلب النابض فيما يعرف بالمكانز اللغوية Corpora التى تعتمد فى عملها على قوة الحاسبات فى المعالجة وقدرات البرمجيات المختلفة فى التحليل والفهرسة والرصد والاسترجاع، وسعات وحدات التخزين فى استيعاب ما لا حصر له من الكلمات والأصوات. فما هى المكانز اللغوية؟

وفقا لموقع جامعة إيسيكس www.essex.ac.uk، فإن هذا المصطلح يشير إلى المجموعات الكبيرة من النصوص أو الملفات الصوتية التى تمثل عينة أو تخصص أو شريحة معينة من اللغة، وهذه النصوص تكون غالبا فى صيغة إلكترونية يستطيع الحاسب قراءتها والبحث فيها، ومن الممكن أن يتكون المكنز اللغوى من نصوص خام فقط بدون أى معلومات أو يحتوى على معلومات لغوية خاصة تسمى بالحواشى أو التلقيب أو الوصف.

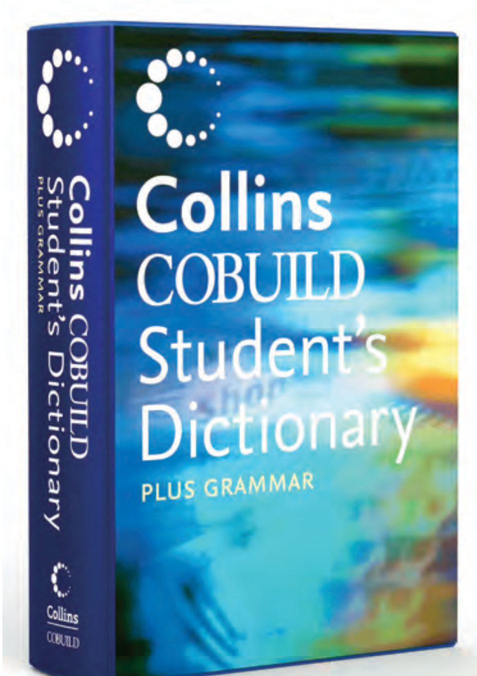
وهى إما مكانز نصية تحتوى على مجموعة هائلة من النصوص المكتوبة والمطبوعة مثل النصوص الكاملة للصحف والمجلات والكتب فى مختلف المجالات مثل الآداب والسياسة والعلوم والفنون أو مكانز صوتية منطوقة تحتوى على تسجيلات آلاف المحادثات والحوارات والخطب ونشرات الأخبار الإذاعية والبرامج الحوارية والمسلسلات والأفلام وغيرها.

مجالات استخدام المكانز

للمكانز دور محورى فى بناء القواميس والمعاجم الحديثة وفى التعرف على خصائص اللغة وكذلك حل مشكلاتها المختلفة، وفى بحث قيم بعنوان بناء مدونة عالية للغة العربية من إعداد الدكتور سامح الأنصارى رئيس مركز لغة الشبكات العالمية فى مكتبة الإسكندرية وأستاذ علم اللغة الحاسوبى بآداب الإسكندرية والأستاذ الدكتور مجدى ناجى رئيس قطاع تكنولوجيا المعلومات بمكتبة الاسكندرية والدكتورة نهى عدلى بكلية الهندسة جامعة الإسكندرية، ذكر الباحثون عدة جوانب لأهمية المكانز فى الدراسات اللغوية نذكر أهمها فيما يلى:

أولا: صناعة المعاجم

تساعد المكانز اللغوية على معرفة معلومات غاية فى الأهمية لبناء المعجم وهى متابعة الكلمات الجديدة التى تدخل اللغة وتحديد وقت دخولها، ومعرفة الكلمات الموجودة بالفعل التى اكتسبت معنى جديدا، ونجد أن أغلب قواميس ومعاجم اللغة الإنجليزية تحتوى على تواريخ مفصلة لكل كلمة وأصلها اللغوى ومتى تم نحتها أو استخدامها لأول مرة،



■ أول قاموس لغة إنجليزية تم إعداده باستخدام المكانز اللغوية الالكترونية

وبمساعدة المكانز الالكترونية يستطيع خبراء صناعة المعاجم البحث فى ملايين الجمل والسياقات المختلفة واستدعاء جميع الأمثلة لكلمة معينة لمعرفة استخداماتها والألفاظ التى ترد عادة قبلها أو بعدها من أجل تحديد التعابير الاصطلاحية والمتتاليات اللغوية، الأمر الذى يسهل تعلم اللغة على الأجانب.

ثانيا: المكانز اللغوية وفهم قواعد النحو

يمكن الاستفادة من المكانز فى دراسة الملامح الصرفية والبحث عن السوابق واللواحق المعينة التى تدخل على الكلمة فكلمة (علم) تتعدد معانيها بإضافة سوابق أو لواحق مختلفة لتصبح (علمية، علمتنا، علماء، تعليم، علوم)، بجانب تحديد توزيع الكلمة وموقعها فى الجملة، وهل تأتى قبل الاسم أم بعد الاسم، وقبل الصفة أم بعد الصفة.

ثالثا: المكانز اللغوية وتحديد الدلالة

فى السنوات الأخيرة، ظهر اتجاه جديد يعتمد على استخلاص معنى الكلمة من المكانز اللغوية، وتتنوع المعلومات الدلالية بين الترادف والتضاد إلى علاقات أكثر تعقيدا، ويمكن استخلاص هذه المعلومات بسهولة من المكانز اللغوية، وينبغى التنبيه، انه يشترط لاستخراج هذه المعلومات أن تكون المكانز شاملة وكاملة بقدر المستطاع، حتى لا يتسرب معنى أو استخدام لا توجد أمثلة له فى المكنز.

رابعا: المكانز اللغوية وتحسين تعليم اللغات الأجنبية

من خلال المكانز من المكن تحليل مدى تكرار وشيوع الكلمات ومعرفة تأثير السياق أو الموقف على أسلوب اللغة وهى معلومات مفيدة للغاية فى وضع مناهج تعليم اللغة العربية سواء للطلاب الأجانب أو العرب أنفسهم. أيضا باستخدام المكانز اللغوية، يستطيع واضع المراجع والمواد إنشاء تمرينات تعتمد على أمثلة حقيقية تقدم للطلاب فرصة اكتشاف خصائص استخدام اللغة. وبدلا من الاعتماد على البحث فى معاجم تقليدية قديمة، يستطيع الطلاب البحث بأنفسهم فى برامج المكانز اللغوية من خلال برنامج بحث وإحصاء لغوى متخصص (concordancing) ويكتشفوا بأنفسهم استخدامات اللغة وقواعدها وخصائصها، ويشجع ذلك على استقلال الطلاب فى التوصل إلى نتائج جديدة بدلا من تعليمهم نتائج متوقعة أو معروفة مسبقا.

خامسا: المكانز اللغوية فى مجالات أخرى

وفى علم اللغويات الاجتماعية، ينصب التركيز على تأثير العمر والنوع والطبقة الاجتماعية والمهنة فى استخدام الأفراد للغة، ولا يمكن دراسة ذلك بشكل سريع ودقيق إلا فى المكانز اللغوية التى تفيد أيضا فى دراسة الأسلوب وتأثره بالمقام أو مقتضى الحال، فالكتابات الأدبية تختلف عن الكتابات السياسية، والكتابات التعليمية تختلف عن المحادثات العادية، وهكذا، وفى كل الأحوال، ينبغى أن تتسم المكانز اللغوية بالشمول، حتى تكون النتائج التى يتم استخلاصها من الدراسة دقيقة وواقعية.

المكانز واللغة العربية

لخص الدكتور سامح الانصارى أهمية المكانز اللغوية لبناء معاجم اللغة العربية فى الجوانب التالية:

١- حصر جميع المعاني: من مميزات البحث الإلكتروني فى المكانز اللغوية إمكانية عرض جميع السياقات التى يمكن أن تظهر فيها الكلمة، وبالتالى إمكانية حصر كل المعانى المختلفة لنفس الكلمة حسب السياقات المختلفة، ومن أمثلة ذلك كلمة (قلب) التى تعنى جوهر وقلب يضخ الدماء ووسط أو منتصف.

٢- دراسة مدى شيوع الكلمات: تساهم المكانز اللغوية فى معرفة أكثر الكلمات شيوعا فى اللغة العربية على المستويين المنطوق والمكتوب، وذلك من أجل تعليم هذه الكلمات للأجانب وحتى يتمكنوا من فهم المعاجم العربية، ومما لا شك فيه أن معرفة مستوى شيوع كلمة ما يدل على أهميتها فى كل اللغة المكتوبة أو المنطوقة، وبالتالى ترشد المتعلم إلى معرفة مدى ضرورة تعلمها أو إغفالها من عدمه.

٣ - دراسة التنويعات المعجمية: دراسة تكرار الكلمة كمادة أو كفئة معجمية (اسم، فعل، صفة، غير ذلك)، ومن المكن أن يكون للكلمة الواحدة أكثر من فئة معجمية حسب السياق، مثال ذلك كلمة (عين) التى ترد كاسم وكفعل مع التشديد بمعنى (وظف).



■ د. سامح الانصارى

٤- دراسة استخدام المترادفات: تحتوى اللغة على كلمات عديدة تعتبر مترادفات لبعضها البعض، ومن خلال المكنز اللغوى يستطيع الباحث بسهولة معرفة مترادفات الكلمة ومعدل شيوعها.

٥- شكل الكلمة وفقا لحالتها الإعرابية: يتغير شكل الكلمة وفقا لحالتها الإعرابية (الرفع والنصب والجر)، ويمكن للباحث من خلال المكنز اللغوى معرفة التنويعات فى شكل الكلمة.

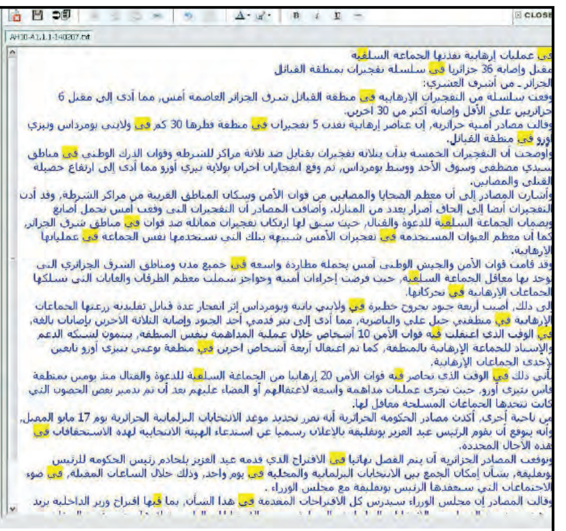
٦- دراسة الكلمة وفقا لاشتقاقها الصرفي: فهناك كلمات ذات أكثر من اشتقاق صرفى مثل كتب (التى يمكن اشتقاق كاتب وكتاب ومكتبة ومكتوب وغيرها)، فالمكنز اللغوى يفيد كثيرا فى هذه الحالات.

٧- معرفة مدى السلامة والصحة اللغوية: أحيانا يكون للكلمة أكثر من جمع أو هى نفسها أكثر من شكل، ويمكن اللجوء إلى ذخيرة المكنز اللغوية الهائلة والبحث فيها لمعرفة أكثر هذه الأشكال شيوعا واستخداما فى اللغة، مثل (شهور وأشهر) و(عيون وأعين)، وهكذا.

٨- معرفة المتصاحبات اللغوية والتعبيرات الاصطلاحية: هناك أوصاف أو كلمات دائما تقترن بأوصاف أو كلمات أخرى دون سبب ظاهر أو منطقي، مثل تعبيرات (فتح الباب على مصراعيه) و(رأب الصدع) و(العروة الوثقى) ودراسة المكانز اللغوية تفيد للغاية فى البحث عن أمثلة هذه المتصاحبات، من أجل تسجيلها فى موادها بالمعجم العربى الحديث. فمثلا فى مادة عروة ينبغى أن يذكر القاموس أنه دائما تليها كلمة (وثقى)، وهكذا.

كيف تم وضع قواعد اللغة العربية منذ أكثر من ألف عام؟ وهل القواعد تسبق الاستخدام أم العكس؟ من المعروف أن الاستخدام يأتى أولا ثم تأتى القاعدة بعد ذلك، الأمر الذى مكن النحاة الأوائل من وضع القواعد بناء على الاستخدام أو الاستعمال اللغوي، فقاعدة رفع الفاعل ونصب المفعول وجر المضاف إليه جاءت من دراسة استخدام هؤلاء الأفذاذ للاستعمالات اللغوية فى أبيات الشعر والحديث والقرآن، وبالمثل يمكن الاستفادة من المكانز اللغوية فى استخلاص القواعد اللغوية على مستوى الكلمة والجملة والخطاب والحصول على معلومات عن تركيب واستخدام العديد من التعبيرات اللغوية.

وقد طرحت الباحثة الدكتورة سلوى حمادة الاستاذة بمعهد بحوث الالكترونيات فى ورقة بحثية بعنوان (نحو منهج عربى مقترح لتصميم المدونات اللغوية) طرق عمل المكانز **١- جمع المادة النصية:** عن طريق جمع المادة الالكترونية المتوفرة على الأقراص الضوئية أو الشبكات أو من خلال الإدخال اليدوى عن طريق لوحة المفاتيح وعن طريق تحويل المادة المنطوقة إلى مادة مكتوبة إن وجدت عن طريق المحولات الصوتية وتحويل البيانات لصورة يسهل التعامل معها آليا مثل صيغة النصوص text أو عن طريق جهاز المسح الضوئى الذى يحول النصوص فى المستندات الورقية إلى الصيغة الإلكترونية.



■ نموذج لمكنز عربي

٢- تجهيز المادة اللغوية وهى مرحلة ما قبل المعالجة: يقصد بذلك عزل الجمل وتقسيم الجمل وعزل الصور وتوحيد الخطوط وحجم الخط وخلافه وعزل علاقات الترقيم وتمييز اللبس والاختصارات والقوائم والشروط بأنواعها.

٣- إدخال معلومات الوصف والترميز والتلقيب المناسبة: يقصد التلقيب annotation أو الوصف meta-dataأو الترميز markup تحديد المعلومات اللغوية وبيانات المؤلف وتاريخ الإنشاء والعنوان واللغة والمجال وإدخال المعلومات اللغوية مثل فئات أو أقسام الكلام والمشتقات والمعلومات النحوية وبناء الجملة والمعلومات الدلالية ومعلومات عن الأسلوب والصوت بصيغة يفهمها برنامج البحث فى المكنز، وسائر أدوات معالجة اللغة، وتتضمن معلومات الوصف تمييز حدود الجمل والعبارات والفقرات وفك لبس حدود الجمل من خلال النقطة والمسافة وتمييز الكلمة الجذر Lemma ومشتقاتها.

عند المعالجة الآلية للغة العربية

٨ خواص لغوية بعضها يعقد الأمر والبعض الآخر يجعله سهلا

بعد كل ما سبق عن المعالجة الآلية تعريفا وتطبيقات ومكونات نصل الآن إلى التساؤل المهم الذى يدور فى أذهاننا جميعا وهو:ما موقف اللغة العربية من المعالجة الآلية؟:هل هى صالحة من الأساس لهذه المعالجة ؟
هل خصائصها تسهل المعالجة الآلية أم تجعلها صعبة؟ وهل حقا تسبب طبيعة اللغة العربية صعوبة فى معالجتها آليا وتحول دون تفاعل الحاسب معها؟ ثم ما أهمية المعالجة الآلية للغة العربية؟ هل هى فى أزمة وتحتاج إلى إمكانيات الحاسب وتكنولوجيا المعلومات لكى تخرج منها ؟ وما هو شكل التفاعل بين اللغة العربية وتكنولوجيا المعلومات؟ فى السطور التالية سنحاول الإجابة.

يقول الدكتور نبيل على أن اللغة العربية أشد حاجة من غيرها إلى هندسة اللغة والمعالجة الآلية بسبب كثرة الفجوات فى تنظيرها اللغوى الراهن، حيث تستطيع المعالجة الآلية بأساليبها العملية التجريبية سد جزء من هذا الفراغ، وإذا لم يحدث ذلك، سيطول الوقت انتظارا لاكتمال الأسس النظرية لمعالجة اللغة العربية الآلية، فهندسة اللغة العربية وتطوير التنظير لها لابد أن يسيرا جنبا إلى جنب لأن كل منهما يتغذى على نتاج الآخر، وحسب رأيه فإن المعالجة الآلية للغة العربية بواسطة الحاسب تنقسم إلى جانبين أساسيين: الأول يشمل نظم البرمجة المستخدمة فى المعالجة الآلية بواسطة الحاسب للفروع اللغوية المختلفة مثل قواعد البيانات المعجمية والذخائر اللغوية والقواميس الإلكترونية ونظام الصرف الآلى الذى يقوم بتحليل الكلمات إلى عناصرها الاشتقاقية والتصريفية أو يعيد تركيبها من هذه العناصر ونظام الإعراب الآلى الذى يقوم بإعراب الجمل آليا ونظام التحليل الدلالى الآلى الذى يستخلص معانى الكلمات استنادا إلى سياقها ويحدد معانى الجمل استنادا إلى ما يسبقها وما يتبعها من جمل.

والجانب الثانى يتضمن التطبيقات التى تقوم على النظم اللغوية الآلية والتى تشمل على سبيل المثال الترجمة الآلية والتدقيق الهجائى والنحوى والفهرسة والاستخلاص الآلى والبحث العميق داخل مضمون النصوص وفهم الكلام ونطقه آليا.

وبالنسبة لخصائص اللغة العربية وعلاقتها بالمعالجة العربية يقول نبيل على فى كتاب العرب وعصر المعلومات ١٩٩٤ أن منظومة اللغة العربية تتصف بمجموعة من الخصائص التى إما تسهم فى تسهيل أو تعقيد معالجتها آليا بشكل يتفاوت من خاصية لأخرى، ويشرح الدكتور نبيل على هذا الأمر بالنسبة لكل خاصية من خواص اللغة العربية على حدة على النحو التالى:

١ـ خاصية التوسط، والمقصود بها الظواهر اللغوية مثل الأبجدية وحالات الإعراب وترتيب الكلمات والتطابق بين الفعل والفاعل وهذا يعنى أن أمور معالجتها آليا ستتعرض للعديد من القضايا، ويمكن لخبراء العرب الاستفادة من الحصاد الهائل المتوافر حاليا من نظم



■ د. نبيل على

معالجة اللغات الطبيعية للانتقاء والتوفيق منها ما يهم شئون لغتنا الوسطى.

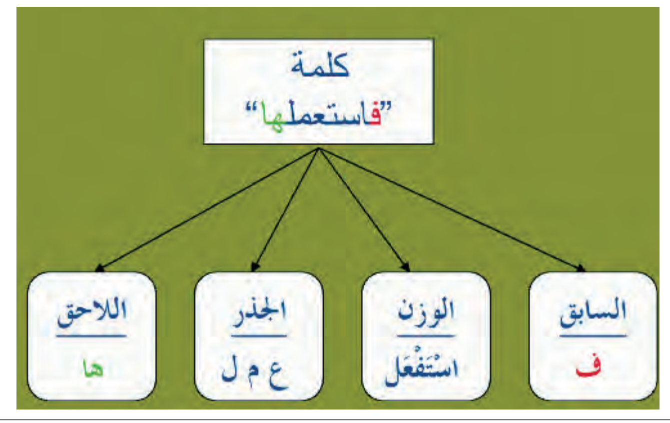
٢ـ خاصية الاشتقاق الصرفى الغزير وهو مع غزارته فإنه شبه منتظم، وهذا الانتظام يجعل اللغة العربية مؤهلة للمعالجة الآلية، وفى هذه الحالة يمثل معالج الصرف الآلى القوم الأساسى فى ميكنة المعجم العربى وتطوير نظم آلية للإعراب الآلى والتشكيل التلقائى.

٣ـ خاصية المرونة النحوية والتى تمثل تحديا حقيقيا للتنظير العربى ومعالجة النحو العربى آليا، ومصدر الصعوبة أن النماذج اللغوية الحالية لأغراض المعالجة الآلية صممت لقواعد النحو الإنجليزية التى تتسم بالصرامة النسبية فى ترتيب الكلمات داخل الجمل، مما يعنى أننا فى حاجة إلى بحوث أساسية وتطبيقية لكتابة قواعد النحو الصورى اللازم لمعالجة النحو العربى آليا يراعى فى كتابته جميع البدائل المكنة لأنماط الجملة العربية بفعل عمليات التقديم والتأخير والحذف والإبدال والإضمار، والبحوث التى أجراها نبيل على توصل إلى أن عدد قواعد الجمل العربية قد يصل إلى ١٢ ألف قاعدة.

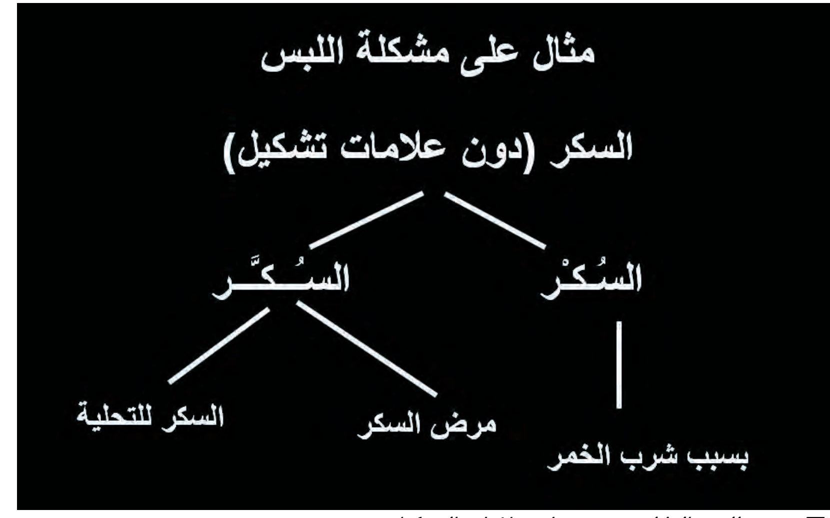
٤ـ خاصية الانتظام فى القواعد الصوتية العربية، حيث يتسم نظام مقاطعها الصوتية ونبرها بالبساطة، حيث لابد أن تجمع المقاطع بحرف صامت ولا تتضمن أكثر من صامتين، وهذه الخاصية ذات أهمية كبرى فى توليد الكلام العربى Speech Synthesis وتمييزه والتعرف عليه آليا Speech Recognition بحيث يبدو الصوت المولد طبيعيا لا ميكانيكيا.

٥ـ خاصية الحساسية السياقية العالية مقارنة باللغة الإنجليزية فشكل الحروف يعتمد على الحرف السابق واللاحق وعلى مستوى النحو تتمثل هذه الحساسية فى علاقات المطابقة مثل تطابق الصفة مع الموصوف والفعل مع الفاعل، مما يجعل معالجة اللغة العربية آليا أكثر تعقيدا وصعوبة من معالجة اللغة الإنجليزية، ويؤدى إلى صعوبات كثيرة فى معالجة النحو العربى آليا.

٦ ـ خاصية غياب تشكيل الكلمات إما بصورة كاملة أو جزئية، والتى تؤدى إلى ظهور حالات معقدة من اللبس، نظرا لتعدد احتمالات قراءة الكلمة مثل كلمة وجد التى يمكن قراءتها لتعنى (عثر على) باعتبار الواو جزء من الكلمة أو لتعنى (واجتهد) باعتبار الواو حرف عطف. ويمثل غياب التشكيل عقبة أساسية أمام



■ نموذج للتفكيك الصرفى فى اللغة العربية



■ نموذج للبس الدلالى بسبب غياب علامات التشكيل

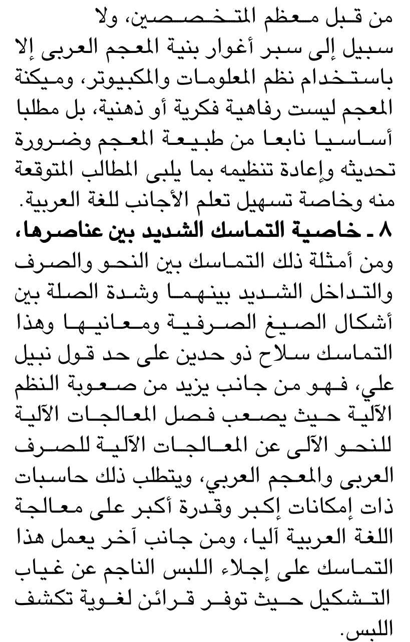
معالجة النصوص العربية آليا، ولهذا تحتاج معالجة النصوص العربية آليا إلى وسيلة برمجية لتشكيل النصوص تلقائيا، وقد نجح نبيل على فى تطوير نظام آلى لإعراب النصوص العربية وتشكيلها آليا وهناك الكثير من البرمجيات التى طورتها شركات عديدة تقوم بإعراب وتشكيل النصوص تلقائيا.

٧ - خاصية اعتماد المعجم العربى على الجذور وليس على الترتيب الأبجدى للكلمات، ورغم صغر نواة المعجم العربى (اقل من ١٠ آلاف جذر) تتعدد المفردات بصورة هائلة بفضل خاصية الاشتقاق الصرفي، وما زالت بنية المعجم العربى المعقدة دون دراسة أو بحث كافية من قبل معظم المتخصصين، ولا سبيل إلى سبر أغوار بنية المعجم العربى إلا باستخدام نظم المعلومات والمكبيوتر، وميكنة المعجم ليست رفاهية فكرية أو ذهنية، بل مطلبا أساسيا نابعا من طبيعة المعجم وضرورة تحديثه وإعادة تنظيمه بما يلبى المطالب المتوقعة منه وخاصة تسهيل تعلم الأجانب للغة العربية.

٨ - خاصية التماسك الشديد بين عناصرها، ومن أمثلة ذلك التماسك بين النحو والصرف والتداخل الشديد بينهما وشدة الصلة بين أشكال الصيغ الصرفية ومعانيها وهذا التماسك سلاح ذو حدين على حد قول نبيل علي، فهو من جانب يزيد من صعوبة النظم الآلية حيث يصعب فصل المعالجات الآلية للنحو الآلى عن المعالجات الآلية للصرف العربى والمعجم العربي، ويتطلب ذلك حاسبات ذات إمكانات إكبر وقدرة أكبر على معالجة اللغة العربية آليا، ومن جانب آخر يعمل هذا التماسك على إجلاء اللبس الناجم عن غياب التشكيل حيث توفر قرائن لغوية تكشف اللبس.

اللبس والمعالجة الآلية للعربية

فى بحثها المنشور فى المؤتمر الخامس لهندسة اللغة ٢٠٠٥، أوضحت الدكتورة سلوى حمادة الباحثة فى معهد بحوث الإلكترونيات أن اللغة العربية من أوضح اللغات وأدقها ومواطن اللبس فيها محدودة وذلك لثراء اللغة بالمفردات وإمكانية استغلال الاشتقاق فى التعبير عن المعانى، واللبس الذى يحدث فى الغالب يرجع إلى غياب التشكيل وعلامات الترقيم والمرونة فى ترتيب الكلمات، مما يسمح بإمكانية التقديم والتأخير، والسياق فكلمة (يسير) تعنى سهل وتعنى يمشي، والسياق فقط هو الذى يحدد هذا المعنى. وذكرت الباحثة عدة مصادر للبس فى اللغة العربية منها التطور واكتساب المعانى الجديدة والتحول الدلالى حيث تستخدم الكلمة بمعنى فى اللغة التراثية وتكتسب معنى جديد أو مغاير تماما فى اللغة الحديثة الدارجة مثل كلمة (جريدة) التى كانت تعنى سعف النخل والآن تعنى صحيفة والاستعمال المجازى وظاهرة تعدد المعاني.



■ د. سلوى حمادة

وهناك فرق بين اللبس بالنسبة للقارئ البشرى والقارئ الآلى (برنامج ترجمة آلية أو تلخيص آلى أو برنامج تشكيل آلي)، فالقارئ البشرى لا يعتبر أغلب صور اللبس التى ذكرناها لبسا لقدرته على فهمها بسبب ثقافته وإلمامه بالسياق واستخدامه للمنطق، أما القارئ الآلى فسيعانى من اللبس فى الوحدات اللغوية مزدوجة التفسير أو متعددة المعاني، لذلك ينبغى تحويل القيود اللغوية – النحوية والدلالية – إلى لوغاريتمات تساعده على اختيار المعنى المقصود. واللبس من أهم أسباب تأخر الحاسبات فى فهم اللغات المكتوبة والمنطوقة، وهناك عدة مستويات للبس فى معالجة اللغة العربية آليا منها المستوى المعجمي، عندما تنتمى الكلمة لأكثر من فئة معجمية (فتكون اسما وفعلا وظرفا وصفة) وفى هذه الحالة يسبب اللبس المعجمى لبسا تركيبيا ودلاليا ومثال ذلك كلمة (ساعة) التى تعنى (عندما) وتعنى (فترة زمنية) وتعنى (آلة لقياس الوقت). ثم المستوى الصرفى وخاصة فى الأفعال المتعدية لمفعولين والصيغ التى تعنى اسم فاعل واسم مفعول فى نفس الوقت وصيغ المبنى للمجهول ومستوى المركبات والربط بين عدة وحدات لغوية مثل تغير معنى الفعل حسب حرف الجر الذى يأتى بعده (رغب فى ورغب عن) والحال مثل (ضربت الرجل ضاحكا) فمن الضاحك الفاعل أم المفعول وربط الصفة بالموصوف دلاليا ومستوى الحذف عندما تحذف بعض الوحدات اللغوية وتستبدل بها وحدات أخرى مثل مرجعية الضمائر والإضافة والصلة والعطف مثل (رأيت الرجال والنساء الحوامل)، وذكرت الباحثة مصادر أخرى للبس يضيق المجال بذكرها جميعا.

ومشكلة اللبس فى اللغة مشكلة ليست سهلة أو هينة، وتمثل عصب التحدى الذى يواجه المعالجة الآلية للغة العربية، ولحل مشكلة اللبس ينبغى حصر جميع الوحدات اللغوية (الأفعال وحروف الجر والأسماء والمصادر) اللبسية بجميع صور اللبس ووضع التفسير المناسب لكل سياق، ثم عمل القاعدة التى تحدد القيود المعجمية والنحوية والتركيبية والدلالية اللازمة لفك اللبس.

فك اللبس بتمثيل المعلومات آليا

ينبغى بحث كيفية فك حالات اللبس سواء أدركها الإنسان أم لا من أجل تمثيل الحلول حاسوبيا ليتمكن الحاسب فى يوم ما من فك اللبس ومن ثم تحليل النصوص العربية تحليلا دقيقا، عندما يمكن استخدام إمكانياته لدعم اللغة العربية والخوض فى غمارها. وقد طرحت الباحثة فى بحث آخر قدمته فى المؤتمر السادس لهندسة اللغة عام ٢٠٠٦ رؤيتها لتصميم نظام حاسوبى لفك اللبس اللغوى فى النصوص العربية الحديثة، وتقوم رؤيتها على توظيف وتوصيف القرائن اللغوية فى فهم المعني، ويقصد به تحديد نوعية القرائن المفيدة فى توجيه المعنى وجمع القرائن من مصادرها المختلفة وتوحيد معايير القرائن على نحو يمكن الحاسب من التعامل معها وتغذية الحاسب بتلك القرائن فى صورة هياكل بيانات مترابطة.

وحددت الباحثة خطوات المعالجة الحاسوبية لفك اللبس حيث يبدأ الحاسب بالبحث فى صور تحليل الجملة المعروفة وعندما يتعرض التحليل لأى نوع من أنواع اللبس يختار القيد المناسب الذى يرتبط بالقيد الإحصائى لأعلى نسبة ورود (من خلال البحث فى قاعدة بيانات مثل الجدول السابق أو مكنز لغوى أو قاعدة بيانات لغوية) فإذا فشل يبدأ فى بحث النسبة الأقل فالأقل. وينبغى أن يبدأ الحاسب بالقيد النحوى حيث يمكن أن يرفض الكثير من الحالات مما يقلل نسبة اللبس أو ينهيها ثم يبحث الحاسب فى السياق الداخلى، فإذا فشل يبحث فى السياق الخارجى والعام، وينبغى لكى تنجح هذه الآلية إنشاء معاجم أو قواعد بيانات تحتوى على هياكل بيانات مترابطة مع مراعات القيود الإحصائية التى تحدد نسبة شيوع اللبس والقياس عليها.

بين طغيان الإنجليزية وتربص الصهاينة

إلى أين وصلت جهود المعالجة الآلية للغة العربية؟

يعرف الجميع أن اللغة العربية تعاني من أزمة حادة وضغوط متعددة في ظل ثورة المعلومات، فمن ناحية هناك طغيان حاد للإنجليزيةعلى لغات البرمجة وعلى شفرات تبادل البيانات المصممة أصلا للغة الإنجليزية وعلى اساليب تخزين واسترجاع المعلومات وعلى المعلومات نفسها على الإنترنت، عن أن القسم الأكبر من البرامج مكتوب باللغة الإنجليزية ومعظم الأبحاث والمراجع والدوريات باللغة الإنجليزية، ومن ناحية ثانية هناك تربص واضح من جانب الصهاينة في اسرائيل باللغة العربية حيث لا تكف محاولاتهم لاختطاف أمور هذه اللغة من العرب والاستحواذ على كل ما يخصها في ميدان المعالجة الآلية ومختلف مسارات التفاعل بين التكنولوجيا واللغة، وفي ظل هذه الضغوط والمخاطر العنيفة: ماذا حقق العرب للغتهم في ميدان المعالجة الآلية؟

لقد فرض طغيان الانجليزية على العرب التحدث باللغة العربية والبحث عن المعلومات باللغة الإنجليزية وتوظيفها بالإنجليزية، لأنه بات من السهل تحويل الكلام الإنجليزي على الورق إلى نسخة إلكترونية، بينما لا نتمكن من ذلك في العربية، ولأن هناك تقنيات للتعرض الصوتي على الكلام بالإنجليزية ولا توجد مثيلتها في العربية، وهناك مواقع متقدمة للترجمة الآلية بين اللغات الأوربية، أما الترجمة من وإلى العربية فتكون عادة رديئة وحرفية وأحيانا كثيرة مضحكة.

وفيما يتعلق بالصهاينة نبه الدكتور نبيل علي في كتابه (الثقافة العربية وعصر المعلومات، ٢٠٠١) إلى ضرورة الانتباه لأن إسرائيل تبدي اهتماما كبيرا بتطبيقات المعالجة الآلية للغات الإنسانية بواسطة الحاسب، وعلى رأسها العبرية، ووصل بها الأمر إلى حد أن تقدمت إسرائيل في منظمة الوحدة الأوربية لتطوير نظم الترجمة الآلية من لغات السوق الأوربية المشتركة إلى العربية (وليس العبرية)، و من الخطير للغاية أن تتولى إسرائيل نيابة عنا مهمة معالجة اللغة العربية آليا فعندئذ تكون قد حلت بنا كارثة ثقافية كبرى، وفي هذا السياق نحاول استعراض الجهود التي بذلت على صعيد المعالجة الآلية للغة العربية والنتائج التي حققتها.



قصور شديد في معالجة اللغة العربية آلياً على مستوى الفقرة

من الملاحظ أن الجهود التي تمت في هذا المجال كان أغلبها إما مشروعات تجارية من الشركات الخاصة أو مبادرات من المنظمات الأجنبية والغربية، وللأسف غابت الجهات الرسمية وجامعة الدول العربية ومجامع اللغة العربية تقريبا عن الساحة، وفي كتاب الفجوة الرقمية (عالم المعرفة، أغسطس ٢٠٠٥، ص ٣٥٧-٣٦٠)، يرى الدكتور نبيل علي والدكتورة نادية حجازي، أن نظم معالجة اللغات الطبيعية آليا قد قد قطعت شوطا بعيدا على صعيد اللغة المكتوبة، لكنها لم تحقق نفس التقدم على صعيد اللغة المنطوقة، وقد حققت معالجة اللغة العربية آليًا نجاحا تكنولوجيا واقتصاديا ملحوظا على مستوى الحرف والكلمة والجملة، غير أن هناك قصور شديد في اللغة العربية على مستوى الفقرة، كما أن التوجه العام نحو تطبيق الذكاء الاصطناعي ونظم معالجة المعارف يتطلب الارتقاء بمستوى المعالجة إلى وحدة البناء الرئيسية للمعرفة ألا وهي المفهوم حتى يتسنى فهم النصوص العربية آليا والنفاذ إلى عمق مضمونها.

إنجازات المكانز باللغة العربية

رغم أن اللغة العربية من اللغات العالمية الرسمية في الأمم المتحدة وتحتل المركز الثالث بعد الإنجليزية والفرنسية من حيث عدد الدول الناطقة بها، وتعتبر من خامس اللغات عالمياً من حيث عدد المتحدثين بها، إلا أن الاهتمام العربي نفسه بالمكانز اللغوية مازال ضعيفا، والسبب في هذا عدم اهتمام الدارسين العرب أنفسهم بعملها. ولا يمكن إنكار فضل بعض الدارسين العرب خارج المنطقة العربية وبعض العلماء في الجامعات المحلية الذين ينحتون في الصخر في هذا الصدد، ويمكن الاطلاع على جدول يحتوي على قائمة بأشهر المكانز الموجودة باللغة العربية في الرابط:

www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm

وللأسف معظمها من إنشاء جامعات وجهات بحثية أوربية وأجنبية. وفيما يلي نبذة سريعة عن أشهرها:

المكنز الدولي للغة العربية

بدأت مكتبة الإسكندرية التي تقوم بدور ملحوظ في نشر الثقافة والمعرفة ودعم البحث العلمي في مشروع كبير لبناء المكنز الدولي للغة العربية International Corpus of Arabic، وهي محاولة حقيقية لبناء مكنز عربي حديث ومتطور يحتوي على ١٠٠ مليون كلمة، وقد تم تجميع العينات المكتوبة للغة العربية المعاصرة من مصادر متعددة روعي فيها أن تكون ممثلة لقطاع إقليمي كبير من الدول الناطقة باللغة العربية وعاكسة بشكل حقيقي وواقعي لأنماط استخدام اللغة العربية في أنحاء العالم العربي.

والمصادر التي اعتمد عليها المكنز الدولي للغة العربية هي الصحف والمجلات العامة والمتخصصة والصحافة الإلكترونية ومقالات الإنترنت والكتب والمصادر الأكاديمية. كما تضمن عدد كبير من الموضوعات مثل العلوم الاستراتيجية والاجتماعية والرياضيات والدين والأدب والإنسانيات والعلوم الطبيعية والتطبيقية من أجل تحقيق شروط التمثيل والانتشار والتوازن في الحجم والمجالات والمصادر.

مكانز نيملار

يهدف مشروع شبكة موارد اللغات اليوروموتوسطية المعروف باسم نيملار NEMLAR موقعه على الويب: www.nemlar.org إلى إنشاء شبكة من الشركاء المؤهلين من منطقة حوض البحر المتوسط لدعم تطوير ذخائر وموارد لغوية عالية الجودة للغة العربية وغيرها من اللغات المحلية بطريقة منهجية معيارية، وتضم الذخائر اللغوية العربية في مشروع نيملار ثلاث ذخائر أساسية هي مكنز نيملار للعربية المكتوبة ومكنز نيملار للكلام

المنطوق والأخبار والإذاعة العربية ومكنز نيملار لتوليف الكلام العربي، وجميع هذه الذخائر مملوكة لاتحاد نيملار. يتكون مكنز نيملار للعربية المكتوبة من ٥٠٠ ألف كلمة فقط من النصوص العربية المعيارية التى تم تصنيفها فى ١٣ مجالا مختلفا (أخبار سياسية، نصوص إسلامية، عبارات شائعة، نصوص من نشرات الأخبار والأدب العربى والأخبار العامة والصحافة العلمية والصحافة الرياضية والنصوص القانونية وشروح مواد المعاجم)، وتهدف إلى إنشاء مكنز متوازن يقدم تمثيلا حقيقيا وواقعيا للتنوع فى الخصائص الدلالية والتركيبية والسياقية للغة العربية الحديثة، وتغطى هذه النصوص فترة زمنية من ١٩٩٠ إلى ٢٠٠٥ . يتكون مكنز نيملار للكلام والأخبار والإذاعة العربية من نصوص ٤٠ ساعة مسجلة من أربع محطات إذاعية مختلفة، أما مكنز نيملار لتوليف الكلام العربى فقد تم إنتاجه للمساعدة فى بناء أنظمة تحويل النصوص إلى كلام، منطوق.

مكنز العربية الفصحى

Classical Arabic Corpus

قام بتجميع هذا المكنز عبد الحميد العوا فى جامعة مانشستر البريطانية، ويحتوى هذا المكنز على ٥ ملايين كلمة ويتكون من نصوص تتضمن أبيات شعرية قصيرة بداية من عصر صدر الإسلام حتى القرن الحادى عشر، وتم الحصول على هذه المادة من الإنترنت، وتنقسم أنواع النصوص إلى أربعة مجالات هى الفكر والمعتقدات والأدب واللسانيات والعلوم.

مكنز العربية العلمية العامة

General Scientific Arabic Corpus

أشرف على إنشاء هذا المكنز أمين المهنى فى جامعة مانشستر البريطانية، وهدف هذا المكنز دراسة صياغة المصطلحات العلمية والتقنية فى اللغة العربية مع التركيز على المصطلحات المركبة، وأخذت المادة اللغوية من موقع مجلة العلم والتقنية الكويتية، وقد تم تلقيب مليون كلمة من هذا المكنز، وتصل دقة التلقيب إلى ٩٢ ٪، ويحتوى على ١٠٠ ألف وصف.

■ نموذج لوظيفة البحث الصرفى فى أحد البرامج التراثية

مكنز العربية المعاصرة المنطوقة Spoken Contemporary Arabic Corpus

يشرف على هذا المكنز الأستاذ دانيال نيومان بجامعة ضرهام البريطانية، وما زال هذا المكنز قيد الإنشاء ويتضمن هذا المكنز عدة مكانز فرعية، منها المكنز العام المنطوق ويتضمن نصوص مقروءة بجانب البث الإذاعى والتليفزيونى والخطب الرسمية والمحاضرات والمقابلات واللقاءات والحوارات والمناقشات والمحادثات العادية والمكنز العام اللا منطوق الذى يتضمن نفس العناصر باستثناء أنها مفرغة ومكتوبة بشكل نصي، وأخيرا المكنز الصوتى (الفونيمي) الذى يضم جملا وأحاديث فى شكل تسلسلى متتابع. ويصل عدد ساعات الكلام فى هذا المكنز إلى ٥٠٠٠ ساعة بما يساوى ٥٠ مليون كلمة.

أشهر برامج المعالجة الآلية للغة العربية

برامج التشكيل والتصحيح الآلي

● **برنامج التشكيل الآلى Arab Diac**: يقوم هذا البرنامج الذى أنتجته شركة آر دى آى بتشكيل النص العربى الخام بنسبة دقة تصل إلى ٩٦٪، وتعد تكنولوجيا التشكيل الآلى الحجر الأساسى لتقنيات الكلام العربيأو تحويل النصوص إلى كلام منطوق.

● **برنامج صخر للتدقيق الإملائى Sakhr Corrector** : يكتشف هذا البرنامج الأخطاء الإملائية العربية والأخطاء الشائعة بالإضافة إلى الأخطاء النحوية.

● **برنامج صخر للتشكيل الآلى Automatic Diacritizer**: لأن اللغة العربية لا تحتوى على علامات تشكيل، طورت شركة صخر برنامج المشكل الآلى الذى يقوم بوضع علامات التشكيل من أجل تسهيل عملية تحويل النصوص إلى حديث منطوق. ويعتمد هذا البرنامج على مستويات متعددة من معالجة اللغة تبدأ من المستوى الصرفى وتنتهى بفك لبس معانى الكلمة استنادا إلى البحث الأساسى المكثف فى مجال معالجة اللغات الطبيعية، وقواعد البيانات اللغوية الضخمة التى طورتها صخر، وقد حقق هذا البرنامج نسبة دقة تصل إلى ٩٨٪.

برامج تلقيب ووصف أجزاء الكلام

يقصد بتلقيب ووصف أجزاء الكلام Part of Speech Tagging تمييز الكلمات فى النص وفقا للوحدات اللغوية المعروفة (اسم، فعل، حرف، صفة، ظرف، حال، إلخ) وفقا لتعريفها وسياقها، بجانب توضيح العلاقة النحوية بين الكلمات المتجاورة فى العبارة أو الجملة أو الفقرة. وهذه البرامج تمثل الخطوة الأولى نحو التشكيل الآلى والإعراب الآلى وتحويل النصوص إلى كلام منطوق.

● **برنامج تلقيب أجزاء الكلام العربىArab Tagger** : تمثل بطاقات التلقيب Tagsلأجزاء الكلام العربى العناصر الأساسية للعديد من العمليات فى معالجة اللغة العربية آليا مثل تحليل التركيب والتشكيل والصرف، وقد صمم هذا البرنامج للتوافق مع الصرف والتركيب العربي، وهى سمة أساسية لهذا البرنامج تميزه عن البرامج المنافسة، وهذا البرنامج من إنتاج شركة آر دى آي.

برامج التحليل الصرفي

وظيفة المحلل الصرفى تحدى جذور الكلمات والمعلومات الصرفية والنحوية وللمحللات الصرفية همية بالغة فى معالجة اللغة الطبيعية، ومشكلة العربية هى غياب علامات التشكيل التى تعد مصدرا أساسيا للبس فى الكلام. ومن المحللات الصرفية:

● **برنامج ArabMorpho** يمثل هذا البرنامج المحرك الأساسى للتحليل الصرفى العربى والوسم أجزاء الكلام والتحليل الدلالى المعجمى العربي. ويحلل هذا البرنامج الوحدات الصرفية والمعجمية باستخدام المقاطع، الأمر الذى يعزز من مرونته ونطاق التغطية فيما بعد ٩٩,٨٪، ويميزه عن البرامج المنافسة التى تحلل الوحدات الصرفية والمعجمية باستخدام ذخيرة الألفاظ، وفى حالة استنفاد القواعد الصرفية، يتم تطبيق تحليل إحصائى ديناميكى من أجل فك غموض الألفاظ بنسبة دقة تصل إلى ٩٦٪. وهو من إنتاج أروى أي.

● **محلل Beesley Xerox**: قام بيزلى بتطوير مولد ومحلل صرفى لاستخدام كأداة مساعدة فى التعليم وكجزء من معالجة اللغات الطبيعية، وهو محلل صرفى من مستويين: المستوى الأول للجذور والقوالب والثانى للزوائد وحروف الجر وأدوات التعريف وأدوات الربط التى ترتبط بالكلمة.

● **برنامج باك وولتر للتحليل الصرفي**: يستخدم من قبل الاتحاد الألمانى للبيانات اللغوية لتدوين النصوص العربية، ويحتوى على ٧٧٨٠٠ جذر تمثل ٤٥ ألف مدخل معجمى ويستخدم المحلل لأغراض الترجمة، لكنه يعانى من خلط العربية بالرومانية ولا يتعرف على علامات الترقيم العربية.

● **برنامج صخر للتحليل الصرفي**: أنتجت شركة صخر المحلل الصرفى الذى يقدم التحليل الصرفى الأساسى للنحو العربى ويغطى اللغة العربية المعاصرة والكلاسيكية. يقوم هذا البرنامج بإزالة الزوائد والمقاطع السابقة واللاحقة من أجل الكشف عن الجذر أو الكلمة.

برامج التنقيب فى النصوص

تعمل هذه البرامج على تحليل النصوص اللغوية من أجل تحديد الأفكار الأساسية والوصول إلى المعلومات القيمة وتسهيل عمليات البحث عن المعلومات، وهذه البرامج مصممة للبحث فى نصوص باللغة العربية وتتميز بقدرتها على فك الكلمات العربية إلى أصولها وجذورها من أجل بحث أكثر كفاءة.

● **برنامج صخر لاستخراج الكلمات المفتاحية** Sakhr Keywords Extractor يحلل هذا البرنامج المستندات والوثائق العربية ويحدد عبارات النصوص وعناصر البيانات الأساسية بشكل آلي. يساعد هذا البرنامج الشركات على تحويل عمليات الأعمال الأساسية من خلال تحديد البيانات الهامة فى كميات كبيرة من النصوص ثم يستخرج أهم التفاصيل لاستخدامها فى المؤسسة أو الشركة. تساعد الكلمات المفتاحية فى تصنيف المستندات فى شجرة موضوعات يحددها المستخدم من أجل سهولة التصفح والوصول للمعلومات، كما يساعد على تشبيك الروابط بين المستندات بطريقة ديناميكية، ويقلل الوقت المطلوب فى البحث عن المحتوى.



● **برنامج محرك البحث النصى العربى** Swift عبارة عن محرك بحث اشتقاقى فى النصوص العربية ويمكن إجراء البحث بالجذر اللغوى أو النمط أو مستوى الكلمة، ويستطيع هذا المحرك التعامل مع استعلامات بحث متعددة المسارات، ويتوافق هذا المحرك مع الويب ونظام الويندوز ولينكس وغيرها من أنظمة التشغيل. البرنامج من إنتاج شركة آر دى آي.

● **برنامج** Arab Dictions يقوم هذا البرنامج بتحليل الكلمات العربية إلى وحداتها الصرفية ومن ثم يتم ربط كل الأشكال الصرفية بمادة المعجم المقابلة لها التى تعتمد على الجذر. يحتوى هذا البرنامج على مواد معجمية ثرية تسهل على القراء خاصة المبتدئين فهم النصوص العربية على كل المستويات. البرنامج من إنتاج شركة آر دى آي.

● **محرك صخر التصنيف سراج:** هذا المحرك يساعدك على تنظيم وتصنيف المعلومات القيمة بكفاءة ودقة فى شجرة موضوعات منطقية أو فئات المحتوى. ويوظف هذا المحرك برنامج التصحيح الإملائى من أجل تصحيح الأخطاء الشائعة فى النص العربى ومستخرج الكلمات المفتاحية، من أجل تحديد القائمة الأولى للكلمات المفتاحية.

● **محرك صخر للتلخيص العربى آليا:** يحدد هذا البرنامج الجمل المهمة آليا داخل النص ويعرضها فى شكل ملخص نصى قصير. يسهل هذا البرنامج عليك قراءة الجمل الهامة فقط فى الوثائق والمستندات الطويلة، ويقلل بذلك من الوقت المطلوب لقراءة المستندات بشكل يدوى.

برامج معالجة الكلام المنطوق

وفقا لشركة جارتنر، ٩٢٪ من كل تفاعلات العملاء تتم عن طريق التليفون، و٨٠٪ من كل المشكلات غير المحسومة عن طريق الويب تنتقل إلى التليفون. فالكلام أو الحديث المباشر هو الواجهة الطبيعية بين الإنسان والحاسب، فالتطبيقات المزودة بالكلام الآلى تفرغ الموظفين للمهام الأكثر تحقيقا للإيرادات. ومن الممكن ربط تقنيات الكلام بتطبيقات الاستجابة الصوتية التفاعلية مثل مراكز الاتصال وأنظمة الرسائل الموحدة والبريد الناطق وأنظمة مساعدة الأدلة وأنظمة الاتصال الصوتى.

● **برنامج** ARAB TALK هذا البرنامج الذى طورته شركة آر دى آى يحول النصوص المكتوبة إلى كلام منطوق ويعتمد على برنامج الشكل الآلى الذى طورته نفس الشركة.

● **برنامج إبصار:** عبارة عن حل متكامل للمعاقين بصريا والمكفوفين، ويعتمد على محركات صخر للتعرف الضوئى على الحروف وتحويل النصوص إلى كلام منطوق. يمكن إبصار المكفوفين من قراءة الكتب والمستندات المطبوعة بالإضافة إلى الملفات الإلكترونية الخاصة بهم دون أى مساعدة من الخارج، ويمكنهم كذلك من كتابة النصوص باللغة العربية والإنجليزية بالإضافة إلى حفظ هذه النصوص وطباعتها بلغة بريل.

● **نظم صخر آلو:** مظلة تجمع كل تطبيقات صخر المزودة بالكلام، وتعتمد على تكنولوجيا التعرف الآلى على الكلام وتحويل النص إلى كلام منطوق التى طورتها شركة صخر وتقوم هذه النظم بتمييز أسماء الموظفين المطلوب الاتصال بهم وبقراءة رسائل البريد الإلكترونى وتحويل الرسائل النصية عبر المحمول إلى كلام منطوق أو العكس.

برامج الترجمة الآلية

طموح العلماء نحو استخدام الحاسب فى الترجمة الآلية كان السبب فى نشوء علوم وتقنيات المعالجة الآلية واللسانيات الحاسوبية، فقد كانت الترجمة الآلية هى أولى التطبيقات المقترحة للحاسبات، وأثبتت التجربة العملية سذاجة التصورات الأولى التى كانت تتخيل القدرة على ترجمة أى نص بضغطة زر، وبعد مرور خمسين عاما على هذه التصورات وبعد كثير من العرق والدموع تمكن الإنسان من تطوير برامج للترجمة الآلية على قدر كبير من الذكاء والجودة.

● **برنامج جهينة:** يعمل هذا البرنامج على مراقبة الأخبار حيث يمسح المئات من مواقع الأخبار العربية والإنجليزية التى تركز على منطقة الشرق الأوسط، ويتيح لك قراءة الأخبار العربية بالإنجليزية، ويستخدم هذا البرنامج محرك صخر للترجمة الآلية ليعرض لك كل المقالات العربية باللغة الإنجليزية.

● **نظام صخر للترجمة للمؤسسات:** عبارة عن حل يمكن المستخدم من ترجمة أى جملة أو فقرة أو مستند من العربية إلى الإنجليزية والعكس عن طريق أحدث تقنيات معالجة اللغات آليا، وهو محرك للترجمة الآلية يعتمد على طريقة التحويل الذى يقوم بتحليل كل المستويات اللغوية تحليلا عميقا بهدف فهم النص المدخل ومن ثم حل كل أنواع الغموض واللبس قبل الترجمة.

مواقع الترجمة الآلية على الإنترنت

لأن الإنترنت بلغاتها العديدة أسمى تجسيد للعولمة على جميع المستويات، ولأن الترجمة الآلية هدفها ترجمة النصوص من لغة لأخرى، كان من الطبيعى أن تظهر مواقع الترجمة الآلية على الإنترنت لتلبية الطلب العالمى الكبير للتواصل ونقل المعرفة.

● **موقع ترجم** www.tarjim.com استخدمت شركة صخر برمجياتها ومحركات القوية فى الترجمة الآلية فى تقديم خدمة الترجمة الآلية عبر الإنترنت للمشتركين فقط، ويستطيع هذا الموقع ترجمة أى نصوص أو صفحات ويب من الإنجليزية إلى العربية أو العكس بسرعة فائقة مع الحفاظ على تصميم صفحات ومواقع الويب.

● **موقع** www.tarjem.com موقع للترجمة الآلية للمواقع من الإنجليزية للعربية والعكس، كما يحتوى على قواميس متعددة للغات.

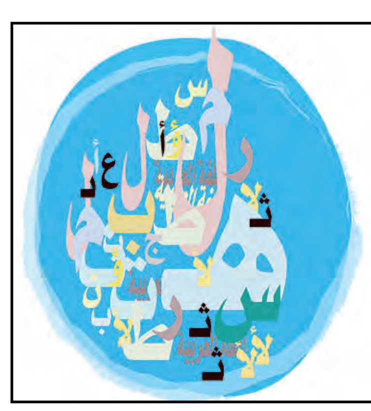
● **خدمة المسبار** www.almisbar.com يتيح لك ترجمة النصوص والمواقع من الإنجليزية للعربية والعكس، ويقدم خيارات لترجمة المختصرات وأسماء الإعلام وإظهار حركات التشكيل.

● **موقع ترجمة جوجل** translate.google.com يعتمد على الترجمة باستخدام الإحصاء ويقدم خدمة الترجمة بين اللغات البلغارية والأسبانية والعربية والإنجليزية والفرنسية والألمانية والسويدية والصينية والروسية والرومانية والبرتغالية والبولندية والنرويجية والكورية والإيطالية والهندية واليونانية.

الخبراء يؤكدون

المعالجة الآلية للغة العربية فى محنة قاسية

يجمع خبراء المعالجة الآلية وهندسة اللغة العرب الذين استطلعنا آراؤهم أو تابعنا بحوثهم ومؤلفاتهم على أن المعالجة الآلية للغة العربية تعيش بالفعل واقعا تعسا تحت ضغوط أزمة عنيفة ومحنة قاسية تهدد وجود ومستقبل اللغة العربية فى عصر العولمة وثقافة الإنترنت وتكنولوجيا المعلومات واقتصاد المعرفة، وهذه الأزمة تجعلها تبدو فى نظر البعض ظلما كما لو كانت غير جديرة بأن تكون لغة العلم والمعرفة، وتتجسد أبرز ملامح هذه المحنة أن الدول العربية والجامعة العربية بلا حضور أو دور ولا تلقى بالا للقضية برمتها، والمحاولات العربية الجارية فى معظمها إن لم تكن كلها محاولات فردية لا تحظى بالدعم، وأغلب المبادرات تتم خارج الوطن العربى فضلا عن غياب الكثير من الأدوات والأسس المطلوبة لدفع جهود المعالجة الآلية للعربية وفى مقدمتها لا غياب معجم عربى حديث يستفيد من المعالجة الآلية وتكنولوجيا المعلومات.





حول رأيه فى الواقع الحالى لتقنيات معالجة اللغة العربية آليا، يقول **الدكتور خالد الغمرى،** أستاذ اللغويات الحاسوبية بكلية الألسن جامعة عين شمس: هناك اهتمام غير مسبوق بمشكلات المعالجة الآلية للغة العربية من جانب العديد من الجامعات والمعاهد والمؤسسات البحثية بالإضافة إلى الشركات- فى العالم كله وعلى وجه الخصوص فى أوروبا وأمريكا الشمالية (الولايات المتحدة وكندا). وقد تمثل هذا الاهتمام فى اتجاهين: الأول تعديل التقنيات المتاحة للمعالجة الآلية للغة الإنجليزية وتطويعها للتعامل مع خصوصيات منظومة اللغة العربية. وهذه المحاولات يقوم بها الباحثون فى الدول غير العربية. أم الاتجاه الثانى فى العالم العربى - فهو تصميم وتنفيذ تقنيات نابعة تماما من اللغة العربية. والملاحظ فى التقنيات الناتجة عن هذه الجهود أنها تهتم بصورة كبيرة بمشكلات الصرف العربى مع بعض الاهتمام بجوانب من المعاجم الاليكترونية. وقد نجحت بعض هذه المحاولات فى حل بعض من مشكلات هذين الفرعين. أما مشكلات النحو والدلالة (المعنى) فما زال الاهتمام بها فى بدايته، ومستويات الدقة فى حل هذه المشكلات غير مشجع. ولابد ألا ننسى أن عمر هذا الاهتمام المكثف بالمعالجة الآلية العربية قصير جدا.

تحديث المعاجم العربية القديمة

وعن رأيه فى الواقع الحالى لصناعة المعاجم العربية؟ يقول الدكتور الغمرى أن هناك بعض المحاولات الجادة لصناعة معاجم عربية تخدم محاولات المعالجة الآلية للغة العربية. فهناك على سبيل المثال - مشروع WordNet أو الشبكة الدلالية للكلمات. وللأسف هو مشروع غير عربى. أعتقد أن ما نحتاجه بحق هو تحديث وتطوير لمعاجمنا القديمة حتى تلبى أمرين أساسيين فى معالجة اللغة العربية: الأول هو تحديثها باضافة ما جد من الكلمات إليها، وإضافة المعانى الجديدة لكلمات قديمة.

>> أغلب المبادرات تتم خارج الوطن العربى

>> لا يوجد معجم عربى حديث يستفيد من المعالجة الآلية

أما الأمر الثانى هو اعادة تصميمها وترتيبها بصوره تسهل استخدامها آليا. ولمجمع اللغة العربية مجموعة من المعاجم المتخصصة المسارد- ثنائية اللغة تنتظر من يستخدمها فى المعالجة الآلية للعربية لبيان نقاط قوتها وأوجه القصور فيها حتى يتم تطويرها لتلبى متطلبات المعالجة الآلية. مشكلة المعاجم عامة أنها تحتاج إلى فريق عمل وتمويل لأنها تستغرق وقتا طويلا وجهدا لايستطيع باحث بمفرده القيام به. بصراحة أنا لا أفهم غياب مشروع عربى كبير لمعجم عربى يلبى متطلبات المعالجة الآلية للغة العربية! كل ما أعرفه جهودا فردية غالبا لاتكتمل فى هذا الاتجاه.

■ د. خالد الغمرى

أما عن كيف تسهم الإنترنت فى تطوير صناعة المعاجم العربية؟ فيقول الدكتور خالد الغمرى هذه الجزئية هى ملخص اهتماماتى البحثية فى معالجة العربية. والاسهامات الممكنة للإنترنت فى هذا المجال لا حدود لها. فهى أكبر قاعدة نصوص عرفتها البشرية، ومتاحة للجميع، وتنمو بصورة رهيبة.

فالحجم الهائل للإنترنت كقاعدة نصوص سوف يسهل بصورة كبيرة جمع معلومات لغوية دقيقة ومتنوعة عن كل كلمة أو تركيب لوضعها فى المعجم. هذا الحجم يتيح معلومات دلالية عن الاستخدامات المختلفة للكلمة فى المجالات المختلفة. الأمر الثانى إثراء المعاجم بكلمات جديدة لم تكن موجودة من قبل. الأمر الثالث توفير معلومات عن التباين فى استخدام الكلمات العربية بين الدول العربية. الأمر الرابع وهذا هو ما أركز عليه هو أن حجم الإنترنت يتيح كما كبيرا مما يعرف بالسياقات الثرية لغويا وهى السياقات التى تتضمن معلومات عن السمات اللغوية المختلفة للكلمة: هناك أيضا العديد من الأبحاث التى استخدمت الإنترنت فى إزالة مايسمى باللبس الدلالي، وهو تحديد المعنى المستخدم فى سياق معين. فكلمة مثل القاعدة قد تعنى تنظيم القاعدة أو قاعدة بيانات أو قاعدة عسكرية أو قاعدة نحوية .. إلخ. فهذه الأبحاث مهتمة بتحديد السياقات المختلفة لكل معنى من هذه المعاني. ومعظم الأبحاث فى هذا المجال تحاول استغلال الطريقة التى ينظم بها



المحتوى على الإنترنت لحل هذه المشكلة.وتوافر مثل هذه المعلومات كما وكيفا سوف يثرى المعجم العربى بصورة كبيرة.

ويضيف الغمرى ان هناك مشكلات وعقبات تعترض صناعة مكانز اللغة العربية وهى وجود هدف وخطة وفريق عمل وتمويل وحل لمشكلات الملكية الفكرية للنصوص التى تتكون منها المكانز. غير إنه هناك محاولات لبناء مكانز عربية فائقة الحجم ولكنها أيضا محاولات غير عربية.

كتب النحو العربى

ويقترح الغمرى عدة نقاط لتعزيز دور تكنولوجيا المعلومات فى خدمة ونشر اللغة العربية والحفاظ على تراثها، وهى إعادة قراءة كتب النحو العربى القديم وتطويعها لمتطلبات المعالجة الآلية. هناك نظريات حديثة كاملة فى علم اللغة أقرأها ملخصة فى سطر أو سطرين من كتاب نحو عربى قديم. الأمر الثانى الاهتمام بالمعاجم العربية القديمة وتحديثها بالطريقة التى ذكرت. هذه المصادر التراثية كنز لغوى يجب أن نحسن استخدامه. اللغة العربية بها من التراث اللغوى مايفيد تكنولوجيا المعلومات. وتكون الخدمة متبادلة. وعن دور مجمع اللغة العربية فى استغلال إمكانيات تكنولوجيا المعلومات فى خدمة اللغة العربية ومعالجتها آليا، يقول الغمرى انه كما ذكرت سابقا هناك على موقع المجمع على الإنترنت مجموعة رائعة من المعاجم والمسارد المتخصصة لابد أن ينتبه إليها الباحثون فى مجال المعالجة الآلية.

محركات البحث

اما رأيه فى كفاءة محركات البحث الحالية باللغة العربية، وما هى نقاط تميزها وضعفها من وجهة نظرك، فيوضح الغمرى ان المشكلة الأساسية فى محركات البحث كلها الآن هو إمكانية البحث بالمعنى لا بالمبنى، أو البحث الدلالى بدلا من البحث الصرفي. أضف إلى هذه المشكلة قضية البحث بالتشكيل فى اللغة العربية. هناك بعض محركات البحث العربية التى تعتمد على محركات بحث مشهورة مثل جوجل Google لا تقدم جديدا، وهناك مجموعة أخرى تظهر بعض الحساسية اللغوية للغة العربية أذكر منها محرك تاياإتTayait وهو محرك بحث عربى يتيح للمستخدم البحث بالأشكال الصرفية المختلفة للكلمة أو مترادفاتها أو مشتقاتها. وهو محرك بحث متميز ورائع. وهناك الحساسية السياسية فى محرك البحث، ولكى تفهم ما أعنيه إبحث عن كلمة مصرباستخدام جوجل مثلا وقارن ترتيب نتائج البحث مع تاياإت. وجرب نفس الشئ مع أية كلمات مرتبطة بعرق أو بنزاع أو قضية عليها خلاف.

الترجمة الآلية

ويرى الدكتور الغمرى أن عملية تطوير إمكانيات الترجمة الآلية من أو إلى العربية، تحتاج الى مزيد من البحث فى علم الترجمة من وإلى العربية والاهتمام بالبعدين الدلالى والثقافى فى الترجمة. فعلى سبيل المثال كلية مثل الألسن عين شمس بها مايزيد عن ١٥ قسما للغات غير العربية لماذا لاتتبنى الكلية مشروعا ضخما عن مشكلات الترجمة بين هذه اللغات واللغة العربية– ثم تتاح نتائج هذا المشروع للمهتمين بالترجمة الآلية لتطويعها واستخدامها فى رفع مستوى الدقة فى الترجمة الآلية. للأسف هناك انطباع سائد بين المترجمين ومدرسى الترجمة أن الترجمة الآلية خاصة من وإلى العربية مشروع فاشل.

تفوق الغرب

وعن السبب فى سر تفوق الجهود البحثية فى مجال الحوسبة اللسانية (أو اللسانيات الحسابية) فى الغرب عموما والولايات المتحدة خصوصا عن الدول العربية، يقول الدكتور الغمرى ان سر تفوق أمريكا فى هذا المجال هو السياسة ثم السياسة ثم السياسة. السياسة الأولى كانت مع بداية الحرب الباردة بين أمريكا والاتحاد السوفييتى فى الخمسينات والستينات من القرن الماضي. فهذه الحرب الباردة كانت السبب الرئيسى وراء ظهور مشروعات الترجمة الآلية بين الانجليزية والروسية. فظن علماء اللغة والحاسب وقتها أن الامر لا يحتاج سوى قاموس بمفردات اللغتين. وكان ذلك أهم خطأ فى تاريخ اللغويات الحاسوبية، فبدأ وعى الباحثين وعلماء اللغة بمدى تعقيد اللغة البشرية وظهرت مجموعة مهمة من النظريات التى تحاول تفسير اللغة منها على سبيل المثال نظرية ناعوم شومسكي. أى أن عمر هذا المجال فى الغرب يصل إلى خمسين عاما. أما السياسة الأخرى فهى سياسة الجامعات والمعاهد البحثية والتى ترتبط بصورة قوية بخطط ونوايا الادارة السياسية. فالمنافسة شديدة بين الجامعات هناك للحصول على التمويل الحكومى للأبحاث



■ محرك تايات يتيح البحث بإستخدام الأشكال الصرفية للكلمات

>> المحاولات العربية فردية لا تحظى بالدعم

توصيات عالم كبير

قدم **الدكتور نبيل على فى كتابه (قضايا عصرية رؤية معلوماتية، دار العين، ٢٠٠٦)** عدة توصيات لاستخدام تكنولوجيا المعلومات والاتصالات فى التعريف بالهوية العربية والتعريف باللغة العربية وإثرائها نوجزها فى السطور التالية:

- استخدام تكنولوجيا الحاسب فى إبراز موقع اللغة العربية داخل أسرة اللغات السامية باستخدام المسارات الزمنية والخرائط اللغوية
- استخدام المعالج الصرفى الآلى فى إبراز خصائص الصرف العربى.
- استخدام تكنولوجيا الحاسب فى بناء أطلس للهجات العربية مع إبراز علاقة هذه اللهجات باللغة العربية الفصحى
- تطوير برامج تعليم ذكية لتعليم وتعلم اللغة العربية باستخدام النظم الآلية لمعالجة الصرف والنحو وقواعد البيانات المعجمية والمعاجم والقواميس الإلكترونية وقواعد وذخائر النصوص العربية والثنائية اللغة.
- استخدام الكمبيوتر فى تطوير معالج نحوى للغة العربية لتفكيك الجمل إلى عناصرها الأولية من



بساطة اللغة الإنجليزية سهلت معالجتها آليا بواسطة الحاسب

<< الدول العربية والجامعة العربية بلا حضور أو دور

فى هذا المجال، هذا التمويل يصل فى حالات كثيرة إلى عشرات الملايين من الدولارات أو يزيد، خاصة بعد حادث ١١ سبتمبر. أما السياسة الثالثة والأخيرة فهى سياسة الشركات التى لها علاقة بهذا المجال والتى ترتبط بعلاقات عمل قوية مع الجهات الأمريكية المرتبطة بالأمن القومي.

وسائل بدائية

وتقول **الدكتورة سلوى الرملى** استاذ الاتصالات بهندسة عين شمس ورئيس جمعية هندسة اللغة،ان هناك نخبة من المتخصصين سواء الفنيين او الادباء ويبذلون جهودا كبيرة فى مجمع اللغة العربية لانشاء المعاجم سواء التراثية او الحديثة، ولكن هذا لا يكفى ولا يواكب العصر، لأنهم يعملون بوسائل بدائية لا تتناسب مع عصر المعلومات الذى نعيشه حاليا، وتضيف الدكتورة الرملى انه يجب على المجمع استخدام تكنولوجيا المعلومات فى عمله لأنها تعطى مساحات واسعة لتخزين البيانات ووسائل متطورة للبحث فى المعاجم المختلفة لاستخراج المعلومات واستخدام الكلمات مع المعاجم الأخرى بصورة سريعة وسليمة.

يتمثل دور جمعية هندسة اللغة فى أن تجعل الناس يستخدمون اللغة العربية فى البحث على الإنترنت والتعامل والتفاعل مع الحاسب والإنترنت باللغة العربية من خلال إدخال واستخراج المعلومات سواء كانت اللغة مكتوبة او منطوقة، حتى يتم الحفاظ عليها وزيادة انتشارها ووقوفها ضد المتغيرات التى تجرى لسيطرة اللغات الأخرى على الإنترنت.

وتضيف الدكتورة الرملى أن هناك بعض محركات البحث باللغة العربية، ولكن محركات البحث حاليا اغلبها باللغة الانجليزية وانه لا يوجد اهتمام من الدول العربية بمعالجة اللغة العربية اليا مع أنها تهدف إلي الحفاظ على اللغة العربية.

وتكمل الدكتورة الرملى انه يوجد فى الدول العربية مكانز للغة العربية ولكنها قليلة جداً وصغيرة جدا فى حجمها وهذا يرجع الى قلة الابحاث فى هذا المجال، ويرجع ايضا الى عدم اتحاد وتعاون الدول العربية لاقامة مشروع او مكنز عربى يضاهى المكانز الاجنبية، وتتمنى الدكتورة الرملى ان يتحد العرب لانشاء مكنز قومى للغة العربية يكون لكن العرب فى كل مكان.

المعجم واللغة العربية

يقول الدكتور **محسن رشوان**، الأستاذ بكلية الهندسة جامعة القاهرة، أنه حتى الآن لا يوجد معجم عربى مجهز لخدمة الحاسب بمعنى أن يكون لدينا معجم عربى رقمى بالكامل بصورة تمكن الحاسب من الوصول إلى الكلمات ومعانيها المختلفة..إننا فى حاجة إلى معجم آلى تفاعلى شامل للغة العربية وليس مجرد معجم أصم يحتوى فقط على الكلمات ومعانيها وتعريفاتها، ولننظر إلى القواميس والمعاجم الإنجليزية، فكلها آلية ورقمية بالكامل مثل قاموس لونجمان وقاموس كوبيلد وأكسفورد وكامبريدج وغيرها.

علماء اللغة لا يعرفون الحاسب

يضيف الدكتور رشوان أن اللغة العربية بطبيعتها لغة تحتاج إلى مثابرة ومجهود كبير من أجل تطويعها للمعالجة الآلية بواسطة الكمبيوتر لثرائها الشديد جدا وتركيبها النحوى والصرفى، بعكس الإنجليزية التى تعد أبسط اللغات الجرمانية، فالعربية أجمل فى اللغة وأصعب فى التحليل، المتاح من الأساليب والتقنيات للغة الإنجليزية لا يكفى حل مشكلات اللغة العربية، ويتطلب ذلك مزيدا من الجهد من قبل العلماء العرب والمشكلة هى عدم توافر التخصص، فنجد مثلا الكثيرين ممن لديهم خلفية فى الحاسب ليس لديهم معرفة باللغة العربية وعلماء اللغة ليس لديهم معرفة بالحاسب إلا قليلا، لذلك يجب ان يقوم اللغويون بدراسة علوم الحاسب والرياضة والتحليل والإحصاء بعمق أكبر.

غياب الاهتمام الرسمى بالمعالجة الآلية

وعبر الدكتور رشوان عن أسفه لعدم اهتمام المؤسسات الحكومية والقومية بمجال المعالجة الآلية، واقتصار النشاط فى هذا المضمار على الشركات الخاصة وتطوعات الباحثين الشخصية، فللأسف أغلب مشروعات المعالجة الآلية وهندسة اللغة العربية تتم بتمويل من الاتحاد الأوربى، وكانت الجامعة والدول العربية أولى بهذا الشرف من أجل إنقاذ اللغة العربية على أيدى أبنائها وليس الأجانب. أتمنى أن يكون هناك تعاون بين جامعة الدول العربية والجامعات المصرية والعربية والشركات تحت إشراف مجمع اللغة العربية حتى نستطيع عمل قواعد بيانات معجمية متكاملة ومكانز شاملة.

أحمد عامر عبدالله

أفعال وأسماء وأشباه جمل وظروف، ويمثل ذلك تحديا كبيرا على مستوى اللغة أو الحاسب. وهذا المعالج النحوى الآلى سيساعد على تحليل مضمون النصوص وفهمها آليا، كما سيسهل عمليات الاستخلاص والتلخيص والفهرسة الآلية ونظم الترجمة الآلية بجانب اكتشاف الأخطاء الهجائية والنحوية بالإضافة إلى حل مشكلة إسقاط علامات التشكيل

● تقديم المعجم العربى على الإنترنت ليس كقائمة من المفردات بل كشبكة مركبة من العلاقات التى تربط بين مشتقات الجذور وصيغ الأفراد والجمع والمترادفات والمتضادات والمصطلحات.

● اللحاق بالموجة الثانية لمعالجة اللغات الطبيعية آليا وتطوير آلة استنتاج عربية لدعم نظم الفهم الآلى وتحقيق متطلبات الويب الدلالية وتطوير آلة بحث ذكية للغة العربية والتوسع فى تطوير النظم الآلية للفهرسة والاستخلاص والتلخيص وتطوير نظم ذكية لتصنيف الوثائق العربية آلياً على أساس المضمون وتطوير نظم التعرف على الكلام العربى بدمج نظم معالجة الصوتيات مع النظم الذكية لمعالجة اللغة آلياً.